CVK-H03341-9-0335083

AMERICAN PHILOSOPHICAL QUARTERLY

⑨

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Arthur C. Danto
James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Hugues Leblanc
Ernan McMullin
Gerald J. Massey
Benson Mates
John A. Passmore
Richard H. Popkin
Richard Rorty
George A. Schrader
Michael Scriven
Wilfrid Sellars
Alexander Sesonske
Manley H. Thompson, Jr.
John W. Yolton

VOLUME 9/NUMBER 1 JANUARY, 1972

CONTENTS

I. STEPHAN KÖRNER: *On the Coherence of Factual Beliefs and Practical Attitudes* 1
II. DAVID S. SCARROW: *Descartes on His Substance and His Essence* . . 18
III. DOUGLAS ODEGARD: *Identity through Time* 29
IV. NICHOLAS LA PARA: *Semantics for Logical and Nomic Modalities* . 39
V. NATHAN ROTENSTREICH: *The Ontological Status of History* . . 49
VI. JANE R. MARTIN: *Basic Actions and Simple Actions* 59
VII. DAVID E. COOPER: *Meaning and Illocutions* 69
VIII. PETER HACKER: *Are Transcendental Arguments a Version of Verificationism* 78
IX. M. B. E. SMITH: *Indifference and Moral Acceptance* . . . 86
X. DALLAS WILLARD: *The Paradox of Logical Psychologism: Husserl's Way out* 94
XI. JAY F. ROSENBERG: *What's Happening in Philosophy of Language Today —A Metaphysician's-Eye View* . 101
XII. SAM COVAL AND D. D. TODD: *Adjusters and Sense-Data* . . . 107
XIII. ROLF GEORGE: *Enthymematic Consequence* 113
Books Received 117

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

V8 /
N2

POLICY

The *American Philosophical Quarterly* welcomes ar [illegible] hilosophers of any country on any aspect of philosophy, substantive or historical. However, [illegible] ficient articles will be published, and not news items, book reviews, critical notices, or "disc [illegible]."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost $12.50, increasing by $1.20 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost $10.10, increasing by $1.20 for each additional 4 pages. Covers will be provided for offprints at a cost of $6 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

335083

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Robert Butts
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Arthur C. Danto
James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
R. M. Hare
Rom Harré
Carl G. Hempel
Raymond Klibansky
Hugues Leblanc
Ernan McMullin
Gerald J. Massey
Richard Rorty
George A. Schrader
Wilfrid Sellars
Alexander Sesonske
Richard Taylor
Manley H. Thompson, Jr.
John W. Yolton

VOLUME 9 (1972)

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 9 (1972)

Allen, Diogenes : *Deliberation and the Regularity of Behavior* . . . 251
Aune, Bruce : *On An Analytic-Synthetic Distinction* . . . 235
Brody, Baruch : *Locke on the Identity of Persons* . . . 327
Castañeda, Hector-Neri : *Intentions and Intending* . . . 139
Cooper, David : *Meaning and Illocutions* . . . 69
Corrigenda . . . 359
Coval, Sam and Todd, D. D. : *Adjusters and Sense-Data* . . . 107
Cummins, Robert and Gottlieb, Dale : *On An Argument for Truth-Functionality* . . . 265
George, Rolf : *Enthymematic Consequence* . . . 113
Gottlieb, Dale, see Cummins, Robert and Gottlieb, Dale
Hacker, Peter : *Are Transcendental Arguments a Version of Verificationism?* . . . 78
Hacking, Ian : *The Logic of Pascal's Wager* . . . 186
Hyslop, A. and Jackson, F. C. : *The Analogical Inference to Other Minds* . . . 168
Jackson, F. C. and Hyslop, A., see Hyslop, A. and Jackson, F. C.
Jones, O. R. : *After-Images* . . . 150
Kekes, John : *Fallibilism and Rationality* . . . 301
Körner, Stephan : *On Coherence of Factual Beliefs and Practical Attitudes* . . . 1
Kupperman, Joel J. : *Aesthetic Value* . . . 259
Lachs, John : *Fichte's Idealism* . . . 311
Ladenson, Robert F. : *Legitimate Authority* . . . 335
LaPara, Nicholas : *Semantics for Logical and Nomic Modalities* . . . 39
Leblanc, Hugues : *Wittgenstein and the Truth-Functionality Thesis* . . . 271
Leslie, John : *Ethically Required Existence* . . . 215
Loux, Michael J. : *Recent Work in Ontology* . . . 119
Machina, Kenton F. : *Vague Predicates* . . . 225
Martin, Jane R. : *Basic Actions and Simple Actions* . . . 59
Martin, Raymond : *On Weighting Causes* . . . 291
Maynard, Patrick : *Depiction, Vision and Convention* . . . 243
McKee, P. L. : *Non-Conscious Seeing* . . . 319
McKinnon, Alastair : *"Existence" in "The Existence of God"* . . . 349
Miller, Peter : *What Aristotle Should Have Said : An Experiment in Metaphysics* . . . 207
Nelson, Jack : *Logically Necessary and Sufficient Conditions for Identity Through Time* . . . 177
Odegard, Douglas : *Identity Through Time* . . . 29
Paulson, Stanley L. : *Two Types of Motive-Explanation* . . . 193
Quest, Edward : *Why Is It Difficult To Deserve Praise?* . . . 343
Rankin, K. W. : *The Non-Causal Self-Fulfillment of Intention* . . . 279
Rosenberg, Jay F. : *What's Happening in Philosophy of Language Today—A Metaphysician's Eye View* . . . 101
Rotenstreich, Nathan : *The Ontological Status of History* . . . 49
Scarrow, David S. : *Descartes on His Substance and His Essence* . . . 18
Sherry, Patrick : *Is Religion a "Form of Life"?* . . . 159
Smith, Malcolm : *Indifference and Moral Acceptance* . . . 86
Snare, Frank : *The Concept of Property* . . . 200
Snyder, A. Aaron : *The Paradox of Determinism* . . . 353
Todd, D. D. and Coval, Sam, see Coval, Sam and Todd, D. D.
Willard, Dallas : *The Paradox of Logical Psychologism : Husserl's Way Out* . . . 94

American Philosophical Quarterly
Volume 9, Number 1, January 1972

I. ON THE COHERENCE OF FACTUAL BELIEFS AND PRACTICAL ATTITUDES

STEPHAN KÖRNER

A CURSORY examination of our factual beliefs and our practical attitudes tends to suggest that they are in important respects similar and interdependent. It seems, in particular, that not only factual beliefs but also practical attitudes may or may not be compatible with each other, may or may not form systematically connected or coherent systems, may or may not be correct—where coherence goes beyond compatibility and correctness beyond coherence. It also seems that the coherence of our practical attitudes presupposes the coherence of our factual beliefs and that the coherence of our factual beliefs in turn depends, though less clearly, on requirements whose acceptance is a matter of practical attitudes. By themselves these observations are, of course, too vague, sketchy, and impressionistic to carry any philosophical weight. They will, however, be clarified, expanded, and justified in the course of the following more detailed inquiry into the coherence of factual beliefs and of practical attitudes.

The argument will proceed as follows. After some brief remarks on factual beliefs, their logical and empirical compatibility or incompatibility (§1), practical compatibility will be defined and different practical attitudes characterized (§2). From practical compatibility there will be distinguished two quite different types of compatibility, namely the concordance of practical attitudes with "reflexive" attitudes directed toward practical attitudes (§3) and the congruity of reflexive attitudes with each other (§4). There follows an examination of the general notion of coherence (§5), of the coherence of factual beliefs (§6), of the coherence of practical attitudes (§7), and of prudence as restricted or qualified practical coherence (§8). A discussion of the relation between the coherence of practical attitudes and their ultimate evaluation (§9) leads to a discussion of their ultimate transpersonal evaluation and of moral coherence (§10) which is shown to coincide with the traditional notion of a morality (§11). Finally the possibility is considered of a relation between the coherence of factual beliefs and an absolute factual reality, and between the coherence of practical attitudes and an absolute moral ideal (§12).

1. Factual Beliefs, their Logical and Empirical Compatibility

A person has a factual belief insofar as he believes that something is the case, i.e., that some entity or set of entities exists objectively or intersubjectively and possesses a certain attribute. When speaking of a factual belief one sometimes means the believing, sometimes that which is believed, and sometimes the linguistic expression of that which is believed. However, such loose talk is rarely harmful since the intended meaning is usually clear from the context and easily put in its proper form. Not all beliefs are factual and not all factual beliefs are correct. Thus the belief that Mr. Pickwick is a generous man is not factual, and the factual belief that Bohemia has a coastline is incorrect. A factual belief is correct insofar as the entitity or set of entities believed to exist objectively or intersubjectively and to possess a certain attribute does so exist and possesses this attribute. Although a person who has a factual belief normally, if perhaps not always, believes it to be correct, he frequently is prepared to admit that it might turn out to be incorrect. By leaving open the question whether correctness is objective or intersubjective, one avoids committing oneself to any particular theory of truth and, in particular, does not prejudge the issue between a coherence- or correspondence-theory of truth.

A factual belief is logically possible if, and only if, the proposition expressing it does not logically imply two contradictory propositions or, what comes to the same, every proposition whatsoever. Two factual beliefs are logically incompatible if the propositions expressing them are logically possible, but their conjunction is logically impossible. Two factual beliefs are logically compatible if they are not logically incompatible. To analyze the notion

C U
CENTRA
LIBRAR

of logical implication, which is a key-notion in the preceding and many of the following definitions, is to formulate a whole system of logic. It will for the sake of definiteness and simplicity be convenient to assume that the persons whose factual beliefs we are examining employ one particular system of logic as fundamental in all their thinking, even if they occasionally make use of other systems in a purely auxiliary function.

We might, for example, assume that their underlying logic is classical elementary logic, intuitionist logic, one of the familiar systems of modal logic, or an extension of these systems which allows not only for exact predicates and definite propositions but also for inexact predicates and indefinite propositions. And there are other possibilities, for example the assumption of a three-valued logic which has been explored by Lukasiewicz and others (Lukasiewicz [1]). The assumption of one of these systems, rather than its alternatives, is not just a technical decision, but may have far-reaching metaphysical consequences concerning such questions as determinism and the freedom of the will. These questions will, however, not be discussed here. (For such a discussion, see Körner [1].) We may, if we wish, assume that the logical implication, which is employed, belongs to classical logic. But the subsequent arguments will remain unimpaired if our metaphysical conscience or conviction demands that it be replaced by the logical implication, say, of intuitionist logic.

What would cause irreparable confusion to these arguments would be to conceive logical implication as an informal relation, accessible to intuition, but incapable of being defined in terms of the principles of any formal logic. By insisting that the logical compatibility or incompatibility of factual beliefs are concepts of formal logic, one does not necessarily condemn as nonsense all kinds of ampliative, substantive, dialectical, and generally informal implications. Indeed even the defenders of such implications, who regard formal logic as a small and insignificant part of the Logic of Reality or of the logical grammar of a natural language, would not deny that factual beliefs may be logically incompatible. Henceforth the notions of logical implication and its cognates, together with the transparent syntax and semantics of the logic to which they belong, will be taken for granted.

In terms of factual belief and logical possibility we can define a notion of empirical possibility which will prove useful in the analysis of practical attitudes. Let us say that a situation or event as described, say, by p is empirically possible for a person, say, A if, and only if, p is logically compatible with A's factual beliefs or, more explicitly, with those of A's factual beliefs which he regards as correct and if A is aware of this compatibility. A person may, of course, not be aware of logical compatibilities and mistakenly believe that logically compatible beliefs are logically incompatible. By excluding these cases in a slightly unrealistic fashion from consideration, one saves oneself a great number of qualifying phrases without any real loss. Two situations or events as described by, say, p and q are empirically incompatible for A if p and q are empirically possible, but their conjunction is empirically impossible, for him. Two situations or events are empirically compatible for A if they are not empirically incompatible for him.

The content and bounds of what for a person is empirically possible vary with his factual beliefs. In a scientific culture it depends largely on the person's scientific beliefs and indirectly on the state of science. It is worth some emphasis that in accordance with our definition each of a pair of logically and empirically incompatible events—for example the occurrence or non-occurrence of a sea-battle tomorrow—might be empirically possible for a person. Whether this seeming bifurcation of the future is merely a matter of insufficient knowledge or whether it is part of a not wholly predetermined reality is one of the metaphysical issues which have been shelved by the logical latitudinarianism which has been adopted for the sake of brevity only.

2. Practical Attitudes, Practical Incompatibility of Situations, Opposition of Practical Attitudes

In contrast to his other likes, dislikes, preferences, and indifferences a person's practical attitudes are directed toward situations or events which he believes to be practically possible. A person's practical possibilities are a species of his empirical possibilities. More precisely, a situation or event as described by, say, p which is empirically possible for a person, say, A is also practically possible for him, if and only if, this empirical possibility includes among the jointly sufficient conditions of the situation's coming into being or of the event's occurring, a deliberate action of A and if A is aware of this. A person may, of course, not be aware of all the jointly sufficient conditions, especially of the part played by a deliberate action of his in making them complete. He may also be mistaken in judging a sufficient condition as insufficient. But we shall again exclude these complications from the discussion to save on the number of qualifications which without advantage would otherwise have to be added. For similar reasons we omit a definition of action. (For such a definition see e.g., Körner [1].)

Two situations or events as described, say, by p and q are practically incompatible for a person A if, and only if, p and q are each practically possible for A but their conjunction is practically impossible for him. Two situations or events are practically compatible for A if, and only if, they are not practically incompatible for him. My spending tomorrow in New York and my spending tomorrow in New Haven are for me two practically incompatible situations or events. Whether or not their separate practical possibility for me is epistemic or ontological, is left just as undecided as the more general question about their separate empirical possibility. It will be convenient to indicate by "$p//q$" that the situations or events described by p and q are practically incompatible and, as the discussion proceeds, to use further symbolic abbreviations. Their purpose is to make the examined relations more transparent and to suggest a more formal treatment.

Practical attitudes are, or are reducible to, preferences or indifferences between practically incompatible situations or events. We shall distinguish between the following: (i) A prefers a situation or event as described by p to a situation or event as described by q with which it is practically incompatible. This can without causing confusion be expressed more briefly by saying that A prefers p to q and by schematically writing $Ap>q$. (ii) A prefers q to p, schematically $Ap<q$. (iii) A is indifferent between p and q, schematically $Ap\sim q$. (iv) A has an (absolute) pro-attitude toward p, schematically $A+p$ or, more explicitly, $Ap>\neg p$. (v) A has an (absolute) anti-attitude toward p, schematically $A-p$ or, more explicitly, $Ap<\neg p$. (vi) A is indifferent to p, schematically $A\pm p$, or, more explicitly, $Ap\sim\neg p$.

The three last mentioned practical attitudes are special cases of the three first mentioned ones. It should, however, be noted that our (absolute) definition of practical pro- and anti-attitudes is not the only definition based on common usage. According to our definition A's pro-attitude toward p means that he prefers p to any q, so long as he is not indifferent between p and q. In this sense one does not, for example, have a practical pro-attitude toward an uneasy truce which one prefers to war and to which one prefers a lasting peace. But there is admittedly another sense of the word in which, in the circumstances of our example, one might be said to have a pro-attitude toward the uneasy truce. Analogous remarks apply to our definition of practical anti-attitudes. The classification of a person's practical attitudes into (absolute) pro-, anti-, and indifferent attitudes is, of course, much cruder than their relative ordering into preferences and indifferences, and is available only if the class of practically incompatible situations or events is finite. It will nevertheless be sufficient for most of our purposes.

Each of the six practical attitudes can be characterized by two features, namely the practical incompatibility between the compared situations or events and the preferred situation or event, if any, as expressed by the corresponding propositions. These are for (i) $Ap>q$ the propositions $p//q$ and p; for (ii) $Ap<q$ the propositions $p//q$ and q; for (iii) $Ap\sim q$ the propositions $p//q$ and the vacuous proposition; for (iv) $A+p$ the propositions $p//\neg p$ and p; for (v) $A-p$ the propositions $p//\neg p$ and p; for (vi) $A\pm p$ the propositions $p//\neg p$ and the vacuous proposition.

Each of these practical attitudes has two opposites. Thus $Ap>q$ has $Ap<q$ and $Ap\sim q$ as its opposites, $Ap<q$ has a $Ap>q$ and $Ap\sim q$ as its opposites, and $Ap\sim q$ has $Ap>q$ and $Ap<q$ as its opposites. Although these relations are quite unproblematic the following general definition of the opposition of practical attitudes will do no harm. Two practical attitudes are opposed to each other if, and only if, (i) the characteristic incompatibility is the same for both of them and if (ii) (a) the preferred proposition is vacuous for one of them and not vacuous for the other or (b) the preferred propositions are different, non-vacuous propositions.

In interpreting $Ap>q$—and *mutatis mutandis* the other practical attitudes—I shall assume what might be called its stability, i.e., that the practical incompatibility of p and q and the preference for p are independent of the wider context in which p and q are compared by A. More precisely, $Ap>q$ is stable if, and only if, for any conjunction of propositions X which is practically compatible with p and for any conjunction Y which is practically compatible with q, A (p-and-X)>(q-and-Y). An unstable preference can sometimes be stabilized by suitably complementing the compared p and q. For example, my present practical preference for soup over orange juice is, as part of my luncheon, stabilized. Some unstable attitudes are, it seems, incapable of stabilization. In particular, one's aesthetic liking or disliking of a situation seems always capable of being turned into either of its opposites by adding further features to it. Indeed it might be reasonably argued that the assumption of stable attitudes is quite generally a more or less unrealistic, though often indispensable, idealization.

To protect the definition of practical incompatibility—and indirectly of practical attitudes—against misunderstandings, two comments are in place. First, it introduces no new "logical relation" —no new logical deducibility—which would justify inferences not justifiable by ordinary formal logic. Thus practical incompatibility between attitudes is defined in terms of practical incompatibility—$p//q$ —between propositions, which is defined in terms of practical impossibility. By stating that p-and-q is practically impossible, one states that another conjunction is *logically* impossible. This latter conjunction contains not only p and q, but many other factual beliefs about the course of nature and about the rôle played in it by deliberate actions. Secondly, the proposed definition of practical incompatibility is, among other things, meant to serve the analysis of practical coherence. Since in most senses of these words incompatibility implies, but is not implied by, incoherence, I have tried to define a narrow sense of practical incompatibility. For, the narrower the sense of incompatibility, the smaller the danger of prejudging the analysis of incoherence.

3. The Concordance and Discordance of Practical with Reflexive Attitudes

Practical attitudes share many of their features with other likes, dislikes, preferences, and indifferences. Among them is the stratification of attitudes into primary attitudes, which are directed toward situations or events not involving attitudes, and attitudes of higher level directed toward attitudes. Because of his stratification a primary attitude and a secondary attitude which is directed toward it may agree or disagree with each other in a characteristic manner. The concordance or discordance of attitudes, as I shall call this type of compatibility or incompatibility, is obviously quite different from practical compatibility or incompatibility which are relations between attitudes of the same level.

It is instructive to compare the stratification of attitudes with the stratification of factual beliefs. A belief, too, may be primary in being directed toward a situation or event not involving a belief, or secondary in being directed toward a belief. Again, just as a person's secondary attitude may or may not be reflexive in being or not being directed toward a primary attitude of this same person, so a person's secondary belief may or may not be reflexive in being or not being directed toward a primary belief of this same person. My dislike of Nero's hatred for all Christians is a non-reflexive secondary attitude, and my belief that he was superstitious a non-reflexive secondary belief. On the other hand, my dislike of my liking for tobacco is a reflexive secondary attitude, my belief that I believe tobacco to be harmful is a reflexive secondary belief. The difference between the stratification of beliefs and of attitudes is most striking if we compare a primary belief and a corresponding reflexive secondary belief with a primary attitude and a corresponding reflexive secondary attitude on the assumption that the two beliefs are concurrent, fully conscious, and untainted by any doubt or wavering and that the name is true of the two attitudes.

If in these conditions a person believes (disbelieves) a proposition p then, whatever p may be, he *cannot* disbelieve (believe) that he believes the proposition. If I, for example, believe (disbelieve) tobacco to be harmful, I *cannot*, in the conditions described, disbelieve (believe) that I believe (disbelieve) tobacco to be harmful. It does not matter whether we construe the "cannot" as expressing a contingent fact about our psychology or a logical consequence of the definition of "belief." The case with primary attitudes and corresponding secondary attitudes is quite different. No factual or logical impossibility attaches to our liking some of our dislikes and disliking others; in disliking some of our likes and liking others; or in changing both our primary and our reflexive secondary attitudes in the course of time. There is nothing absurd in a person's liking ceremonies and liking his liking for them, in a person's liking ceremonies and disliking his liking, or in the other ways in which a person's primary attitude toward ceremonies and his reflexive secondary attitude toward his primary attitude may be in concordance or discordance.

The different relations between primary practical and reflexive secondary attitudes can be classified in a variety of ways of which the following is best suited to our purpose. Let us say that a primary, practical, and a reflexive secondary attitude are in concordance with each other if, and only if, whatever the primary attitude may be, the secondary attitude is a pro-attitude toward it (or, what comes to the same, if the secondary attitude is a preference for the primary attitude over those primary attitudes with which it is practically incompatible). If a primary, practical and a secondary reflexive attitude are not in concordance with each other they are in discordance with each other. We may further distinguish between strong discordance, if the secondary attitude is an anti-attitude and weak discordance, if the secondary attitude is an attitude of indifference. This classification can—with an eye to its ethical relevance—be put more expressively, if we give the names of domination to the relation between reflexive secondary and primary practical attitude, of dominating attitude to the secondary, and of dominated attitude to the primary attitude. We then say that a reflexive pro-attitude positively dominates, that a reflexive anti-attitude negatively dominates and that a reflexive attitude of indifference indifferently dominates the primary attitude toward which it is directed. To a positively, negatively, or indifferently dominating attitude there corresponds, of course, respectively, a positively, negatively, or indifferently dominated attitude.

Examples come easily to mind. Thus a person may have a primary preference for building a garage over the planting of a garden on a site which has no room for both. If his reflexive attitude toward this preference is a pro-attitude we have an example of concordance or positive domination; if an anti-attitude, an example of strong discordance or negative domination; if an attitude of indifference, an example of weak discordance or indifferent domination. Only a wholly self-satisfied person could be wholly unaware of the discordance of some of his attitudes. Yet even he, unless he is also wholly unobservant and unreflective, could be made to notice that, though he himself may be free from such inner conflict, ordinary human beings are frequently exposed to it. When we are aware of a concordance between a primary and a secondary attitude we tend to feel that our secondary attitude supports our primary. When we are aware of a strong discordance we tend to feel that our secondary attitude weakens our primary. Insofar as *some* of our reflexive attitudes are, as we shall see, prudential or moral, the relations of concordance and discordance are likely to prove relevant to the analysis of the conflict between prudence and inclination and between duty and inclination.

It may well be that a person's concordant or discordant attitudes are not primary and secondary, but secondary and tertiary attitudes or attitudes of still higher levels. A person may, for example, have an irradicable dislike of committees and up to now have liked his dislike. He may, however, for a number of reasons, such as growing awareness of the social importance of committees, become first uneasy about his liking of his dislike and later even come to dislike his liking of his dislike of committees. He may, or may not, at the end of his life come to dislike his dislike of committees, even if he dies disliking them. There is, in any case, nothing absurd about the discordance of a tertiary and a secondary attitude. It is in principle not difficult to extend the relations of concordance, discordance, and domination from applying to attitudes of primary and secondary, secondary and tertiary, and quite generally attitudes of immediately neighboring levels, to attitudes of only indirectly connected levels. One might, for example, define an attitude of level higher than two as concordant with a primary practical attitude if, and only if, it is a pro-attitude toward a pro-attitude . . . toward a pro-attitude toward the primary attitude. But I shall disregard these complexities since they do not require any substantially new treatment.

4. THE CONGRUITY AND INCONGRUITY OF REFLEXIVE ATTITUDES

So far we have considered the following kinds of incompatibility and, of course, of the corresponding compatibility: the logical and empirical incompatibility of factual beliefs, the practical incompatibility of (propositions describing) situations or events, the opposition of practical attitudes, and lastly the discordance of practical with reflexive attitudes. All of these appear to be based on intellectual shortcomings, except the discordance of practical with reflexive attitudes, which often strikes one not as a defect but rather as a redeeming feature of humanity. We now turn to relations of incompatibility and compatibility between reflexive attitudes. Their incongruity or congruity, as these relations will be called, are defined on the one hand in terms of the practical incompatibility of situations or events and the opposition of practical attitudes, on the other in terms of the discordance of reflexive and practical attitudes.

A reflexive attitude is a pro-, anti-, or indifferent attitude of a person directed toward a practical pro-, anti-, or indifferent attitude of his own which is in turn directed toward a situation or event. In defining the incongruity of reflexive attitudes it is convenient to start, as it were, at the bottom of their two-tiered structure by considering certain limiting cases of practical attitudes, namely practical attitudes which are incapable of being realized, practical attitudes which are incapable of not being realized and pairs of opposite practical attitudes. A practical attitude is (i) incapable of being realized or, more briefly, absurd if, and only if, (a) it is a pro-attitude toward a practically impossible situation or event, in symbols $A+(p \wedge \bar{p})$ . . . where $p//\bar{p}$ or (b) it is an anti-attitude toward a practically necessary situation or event (a situation or event described by a proposition the negation of which describes a practically impossible situation or event), in symbols $A-(p \vee \bar{p})$ . . . where $p//\bar{p}$ and p and $\bar{p}$ in addition exhaust all empirical possibilities. An example would be A's anti-attitude toward either burning his farm or leaving it to the enemy, where there is no other empirical possibility.

A practical attitude is (ii) incapable of not being realized or, briefly, trivial if, and only if (a) it is a positive attitude toward a practically necessary situation or event, in symbols $A+(p \vee \bar{p})$ . . ., where $p//\bar{p}$ and p and $\bar{p}$ exhaust all empirical possibilities or (b) it is a negative attitude toward a practically impossible situation or event, in symbols $(A-p \wedge \bar{p})$. . . . Our last limiting case is (iii) the case of two opposed practical attitudes, for example when one and the same person is both for and against the realization of a practical possibility, i.e., $(A+p) \wedge (A-p)$. (The opposition of practical attitudes has been defined above).

Corresponding to the three limiting cases of practical attitudes and their possible combinations, we can now define the incongruity of reflexive attitudes. Two reflexive attitudes are incongruous with each other if, and only if, they are not directed toward either absurd or trivial practical attitudes and (i) if they are each in concordance with (if they each positively dominate, are pro-attitudes directed toward) a different practical attitude, such that the compound of these practical attitudes is an absurd attitude; or (ii) they are each in discordance with (if they each negatively dominate, are anti-attitudes directed toward) a different practical attitude such that the compound of these practical attitudes is a trivial attitude; or (iii) they are between them (a) in accordance with a practical attitude and one of its opposites or (b) in discordance with a practical attitude and both of its opposites. Two reflexive attitudes which are not incongruous are congruous with each other.

Somewhat more simply, but less precisely, we might say that two reflexive attitudes are incongruous with each other if together they (i) favor an absurd practical attitude, (ii) disfavor a trivial practical attitude or (iii) (a) favor a practical attitude and one of its opposites or (b) disfavor a practical attitude and both of its opposites. If, as before, $p=A$'s burning of his farm, $\bar{p}=A$'s leaving it to the enemy and if p and $\bar{p}$ are practically incompatible and exhaustive of all empirical possibilities, we can illustrate the various types of incongruity by the following examples: (i) Reflexive attitudes: $A+(A+p)$, $A+(A+\bar{p})$, compound practical attitude: $A+p \wedge \bar{p}$, compound reflexive attitude: $A+(A+p \wedge \bar{p})$, i.e., favoring an absurd practical attitude. (ii) Reflexive attitudes: $A-(A-p)$, $A-(A-\bar{p})$, compound practical attitude: $A-p \wedge \bar{p}$, compound reflexive attitude: $A-(A-p \wedge \bar{p})$, i.e., disfavoring a trivial practical attitude. (iii) (a) Reflexive attitudes: $A+(A+p)$, $A+(A-p)$. Here $A+p$ and $A-p$, i.e., one of its opposites, are favored; (b) Reflexive attitudes $A-(A+p \wedge A\pm p)$, $A-(A-p)$. Here $A-p$ and both of its opposites, i.e., $A+p$ and $A\pm p$ are disfavored.

In order to emphasize the analogy between incongruous attitudes and practically incompatible attitudes, we may define a notion of internal incongruity which, like internal practical consistency or practical possibility, characterizes one attitude rather than a pair of attitudes. We simply call a reflexive attitude internally incongruous if, and only if, it favors an absurd or disfavors a trivial practical attitude, or else favors a practical attitude and one of its opposites or disfavors a practical attitude and both of its opposites. An internally incongrous, reflexive attitude can clearly be decomposed into mutually incongruous reflexive attitudes none of which is internally incongruous, just as a practically impossible attitude can be decomposed into practically possible attitudes.

The congruity or incongruity of reflexive attitudes, as well as the concordance or discordance of practical with reflexive attitudes, have, as far as I can see, never been clearly discerned, because the stratification of attitudes has on the whole been neglected. The so-called attitude-theorists of ethics in particular tended to treat all attitudes as being on the same level and consequently ignored not only congruity and concordance, but also some related systematic connections between attitudes of different levels which together form an important aspect of morality.

5. The General Notion of Coherence

We are now in a position to analyze the systematic connection or coherence of a person's factual beliefs, of his practical attitudes and thereby—because his practical attitudes involve his factual beliefs—of his practical attitudes with his factual beliefs. It seems, however, advisable and ultimately more economical to preface the analysis proper with some preliminary remarks both on the notion of coherence, which is the subject matter of the analysis, and on some other notions of coherence with which it should not be confused. In these preliminary remarks the terms "compatibility" and "incompatibility" will be used in an ambiguous manner, that is to say without distinguishing between their various senses. The distinctions will, of course, be carefully observed in the more detailed examination of factual and practical coherence to which they are essential.

A person's factual beliefs or practical attitudes are systematically connected or coherent insofar as they are not only compatible with each other, but in addition satisfy certain requirements of systematic connection. Any such requirement or principle of coherence is characterized by what may be called its comprehensiveness, its selective force, and its manner of acceptance. By saying that a principle of coherence is comprehensive I mean that, by itself or in conjunction with a finite number of similar requirements, it imposes a condition on all factual beliefs, for example that they should refer to material objects (rather than events, monads, etc.) or on all practical attitudes, for example that their realization should not involve any violation of human life. The selective force of a principle of coherence, which depends on the content of the imposed condition, consists in rejecting from any two incompatible beliefs or practical attitudes, one of which is incompatible with the principle, that belief or attitude which is so incompatible. As to its manner of acceptance, a principle of coherence is accepted in preference to any principle with which it is incompatible in such a way that this preference is not incompatible with some attitude of higher level. If a person accepts a number of principles of coherence, they must, of course, be compatible with each other.

From any specific notion of factual, practical, or factual-cum-practical coherence, as determined by one or more principles of coherence, we must distinguish two other notions—or families of notions—which might be called "intuitive" and "total" coherence. The former is assumed to be an intuitively well understood characteristic of systems of beliefs, attitudes, or beliefs and attitudes, whose instances and counter-instances are clearly recognizable, at least when the systems are properly presented. The notion of total coherence involves the additional assumption that all intuitively coherent systems are sub-systems of one and only one maximal system, that is to say the total coherent system.

The notion of intuitive coherence, as based on common usage[1] or as an undefined technical term is too vague for our purpose. Its employment in philosophical arguments involves no clear indication of its relation to specific notions of coherence, such as causal coherence (that is to say the imposition of a causal connection on the situations or events which a person believes to exist at some time and to possess certain characteristics). The notion of total coherence contains in addition to the vagueness of the notion of intuitive coherence an element of obscurity. Its proponents are wont to assert that any proposition of the maximal system "entails" every proposition of the system and no proposition outside it and that the maximal system contains all true propositions and only such. The concept of entailment employed is left unexplained, except that it is characterized as an intensional relation between meanings and as quite different from the logical implication or deducibility-relation of extensional logic. (For a criticism of this kind of coherence theory see Körner [1].)

To avoid the pitfalls which arise from using the vague notion of intuitive and the obscure notion of total coherence only specific principles of coherence will be considered or, at least, classes of such principles for which specific instances are readily identifiable in contemporary thought or the history of ideas. Among them are, in particular, the principles describing the "categorial framework" of a person's factual beliefs and practical attitudes, since these principles satisfy our general definition of a principle of coherence. Apart from principles of coherence we shall also examine some less comprehensive principles, provided that they present options for systematically organizing a large area of factual beliefs or practical attitudes in one way rather than another. Consideration of such principles of "restricted coherence" will throw some light on the relation between coherence and truth.

---

[1] See, for example, the *Oxford English Dictionary*.

6. On the Coherence of Factual Beliefs

In making the transition from principles of coherence in general to principles of factual coherence—more precisely, coherence of the factual beliefs of a person, say, *A*—one must in the former replace the ambiguous sense of compatibility and incompatibility by the appropriate relevant senses. This is easily done: A principle of factual coherence is comprehensive in that, by itself or in conjunction with a finite number of similar principles, it imposes a condition on all *A*'s factual beliefs. It is selective by eliminating from any two logically incompatible beliefs, one of which is logically incompatible with the principle, the belief which is so incompatible. It is accepted in preference to any principle with which it is logically incompatible, in such a manner that the preference on the one hand is not strongly discordant with (is not negatively dominated by) some attitude of higher level and on the other hand is congruous with any of *A*'s principles of practical coherence.

A person's factual beliefs are all accommodated by what may be called the "categorial framework" of his factual thinking; or, what amounts to the same, they satisfy the "categorial principles" employed in his factual thinking. Since these principles are by virtue of their comprehensiveness, selective force, and manner of acceptance principles of factual coherence, a more detailed analysis of factual coherence can draw heavily on the analysis of categorial frameworks, which ever since Aristotle has been at the center of philosophical interest.[2]

In holding any factual belief, that is to say in believing that some entity or set of entities exists objectively or at least intersubjectively and possesses a certain attribute, one distinguishes between particulars and attributes. One also categorizes all particulars—be they independent substances or dependent on them for their existence—into maximal kinds (*summa genera*). Each maximal kind, for example the class of material objects (of events, of bodies, of minds, of spirits, etc., depending on the person's way of differentiating his experience into particulars and attributes and of classifying them into "natural" kinds) is associated with one or more constitutive principles and, at least some of them, with one or more individuating principles. A constitutive principle expresses a necessary condition for a particular's belonging to a maximal kind; an individuating principle a necessary and sufficient condition for a particular's being a distinct member of maximal kind.

More precisely, if M is a maximal kind and C a non-empty attribute, then the proposition that "Being an M" logically implies "Being a C" expresses a constitutive principle associated with M. Again, if M is a maximal kind, if D is a non-empty attribute, of which D_0 is a minimal sub-attribute (*infima species*, determinate), then the proposition that "Being a distinct individual instance of M" logically implies, and is logically implied by "Being a D_0" expresses an individuating principle associated with M. In accordance with philosophical tradition the non-empty attribute C might be called a "constitutive," and the non-empty attribute D an "individuating" attribute of (the members of) M. For example, in an event-ontology of the kind derived by Whitehead through extrapolation from the general theory of relativity, the class of events is a maximal kind, the causal connectedness (as required by the theory of relativity) of any event with any other a constitutive attribute, their location in a Riemannian space-time continuum the individuating attribute of events. The corresponding constitutive and individuating principles express respectively that the former attribute is a logically necessary condition of an entity's being an event, and that the latter attribute is a logically necessary and sufficient condition of the event's being a distinct individual event.

Apart from the differentiation of all factual beliefs into particulars and attributes, the categorization of particulars into maximal kinds, the association with the maximal kinds of constitutive and individuating principles, a categorial framework also includes a system of logic which, among other things, determines the meaning of the logical implication in terms of which, for example, the constitutive and individuating principles are formulated. The (finite) conjunction of all the propositions which together describe a person's categorial framework, has clearly the character of a principle of factual coherence. But even one of the conjuncts—for example, the second analogy of Kant's *Critique of Pure Reason* which exhibits a certain categorial framework—may have his character. It will here be assumed that as a matter of empirical fact every normal human being employs a categorial framework in his thinking, even if he is not aware of all its features, even if some of its features are rather indeterminate, and even if he may on occasion waver between the acceptance of different categorial principles. It will be further assumed as an empirical fact that not all people employ or have employed the same categorial framework, but that categorial frameworks may differ, and often have differed, from person to person, society to society, and age to age. (For a more detailed discussion of categorial change, see Körner [2].)

---

[2] Having only fairly recently examined the structure and function of categorial frameworks, I can and indeed ought to be content with some brief summary statements. For details see Körner [2].

Principles of factual coherence, which are characteristic of a person's categorial framework, must be distinguished from principles which are accommodated by his categorial framework, but resemble categorial principles in their wide, though not comprehensive, application, in their selective force, and their manner of acceptance. Such restricted, limited, or qualified principles may resemble the principles of factual coherence also by differing from person to person. One of the sources of this divergence are deep-seated and far reaching antinomies, which can be removed in a number of mutually exclusive ways, of which not all seem equally preferable to everybody. A famous example is Russell's antinomy. Another source of the divergence of restricted principles of coherence are non-monomorphic (non-categorical) theories, which can in a number of mutually exclusive ways be made monomorphic. A well-known example is Euclidean geometry from which the parallel-postulate and all its deductive equivalents have been eliminated. It seems worthwhile to emphasize these divergences. For while the divergence of scientific theories and specific common-sense beliefs is generally acknowledged, the divergence of restricted and even more of unrestricted principles of factual coherence is frequently ignored or regarded as an illusion.

7. On the Coherence of Practical Attitudes

In defining the notion of a principle of practical coherence one must once again replace the ambiguous sense of compatibility and incompatibility which occurs in the general definition of a coherence principle by the appropriate relevant senses. To do this one merely needs to observe that a principle of practical coherence imposes a condition on practical attitudes and expresses a preference (a secondary, reflexive preference) for attitudes which satisfy this condition. A principle of practical coherence acknowledged by a person *A* is comprehensive in that, by itself or in conjunction with a finite number of similar principles, it imposes a condition on all practical attitudes. It is selective in that it eliminates from any two practically incompatible attitudes, one of which is discordant with the (secondary attitude expressed by the) principle, the attitude which is so discordant. As to its manner of acceptance, a principle of practical coherence is preferred to any other secondary attitude with which it is incongruous, in such a manner that this preference is not strongly discordant with (not negatively dominated by) any attitude of higher level.

As an example of a possible principle of coherence we might consider the following requirement which is implied by all versions of utilitarianism and is, it seems, widely accepted in our culture: For any two practically incompatible (situations or events described respectively by the) propositions *p* and *q* such that *p* differs from *q* only in involving the happiness of one more person, the preference of *p* over *q* is preferred both to the preference of *q* over *p* and the indifference between them. This principle, which expresses a secondary attitude, is obviously comprehensive and selective. It is accepted in preference to any principle with which it is incongruous, for example the principle that *p* always be preferred to *q*, whenever *p* leads to an increase in the production of commercial commodities. For those who accept the principle, the secondary preference expressed by it is not discordant with any attitude of higher level.

Among everybody's principles of practical coherence is the requirement that all his practical attitudes be directed toward situations or events which he regards as belonging to a factually coherent system. This requirement simply expresses the dependence—implied in the very notion of a practical attitude—of a person's practical attitudes and their practical coherence on his factual beliefs and their factual coherence. It represents one side of the interdependence of practical attitudes and factual beliefs. The other side, as has been pointed out earlier, is represented by the circumstance that every principle of coherence—practical or factual—is the object of a preference, belonging to a system of attitudes which is required to be practically coherent. Thus in a Kantian ontology the principle of causality, which is one of its principles of factual coherence, is accepted as a matter of practical preference and thus subjected to the practical coherence-principles of this ontology; while the categorial imperative, which is one of its principles of practical coherence, in turn satisfies the requirement that the practical attitudes corresponding to it be directed toward a system of situations or events which are factually, and in particular causally, coherent.[3]

Just as we distinguish principles of factual coherence from less comprehensive principles to which they bear some resemblance, so we must also distinguish principles of practical coherence from limited or restricted principles of practical coherence. Such principles may, in particular, be directed toward a class of practical attitudes or they may rest on a theoretical idealization or fiction to the effect that all practical attitudes share some characteristic which in fact they do not possess. Both kinds of restriction are exemplified by the so called principles of prudence. In commonsense thinking these principles are usually regarded as directed toward all practical attitudes which are not immoral—morality being superior to prudence. In economics and related theoretical fields, on the other hand, the principles of prudence are usually construed as the only principles of rational conduct—moral principles being excluded from consideration on heuristic or other grounds.

8. Prudence as Restricted or Qualified Practical Coherence

A person's, say *A*'s, principle of prudence expresses a secondary attitude which (i) is directed toward all of *A*'s morally indifferent practical attitudes or to all his practical attitudes treated *as if* they were morally indifferent, and which (ii) alone or in conjunction with other such principles selects from any set of practical preferences or indifferences the maximal preference or a set of maximal preferences. Among the main reasons for our interest in prudential principles are the moral relevance of prudence, since prudent conduct may be a duty toward others or oneself; the analogy between the divergence of very general factual and the divergence of prudential principles; and last but not least the light thrown by recent decision-theoretical analyses of prudential conduct—however narrowly or abstractly conceived—on the notion of practical coherence. To make proper use of these analyses one has to be rather careful in distinguishing between the analyses of various concepts of prudence and the propaganda for them, as well as between the empirical fact that people tend to behave in certain ways and the other empirical fact that they often accept standards of conduct of which their actual conduct falls short. We are here interested in prudential principles as standards, more precisely as principles of (restricted) coherence.

The principles of maximization suggested by the theoreticians all imply that the person who accepts such a prudential principle is both able and willing to order his practical preferences and indifferences in accordance with the following postulates, in which p, q, r are (propositions describing) situations or events which A believes to be pairwise practically incompatible. Trichotomy: For all p and q—$Ap>q$ or $Ap<q$ or $Ap\sim q$. Transitivity of preference: For all p, q, r—$Ap>q$ and $Aq>r$ together imply $Ap>r$. Transitivity of indifference: For all p, q, r—$Ap\sim q$ and $Aq\sim r$ together imply $Ap\sim r$.

None of these postulates are implied by our definitions of practical attitudes and of their practical compatibility or incompatibility. Thus although the definitions imply that every preference or indifference has two opposites, they do not imply the postulate of trichotomy, i.e., that any two practically possible situations or events are comparable. The possibility that, contrary to Bentham's thesis, playing pushpin and reading poetry may not be comparable, has not been ruled out. Again, the two postulates of transitivity which relate three distinct practical possibilities are not implied by our definitions which relate only two. Indeed one can imagine circumstances in which the transitivity postulates might be "prudently" violated. For example, it might—say in accordance with the maxim *de minimis non curat lex*—be prudent to be indifferent between p and q and between q and r, because they differ only minimally, but not between p and r which differ more than minimally. Similarly, a person whose overriding desire is to appear modest might prudently prefer accepting a gift p (more precisely, p describes the acceptance of the gift by the person) to a slightly smaller gift q, and a gift q to a slightly smaller gift r, because in each of these cases the difference is small enough not to arouse the suspicion of immodesty, while the fear of this suspicion may make him indifferent between p and r between which the difference is big enough to arouse suspicion.

Insofar as a person accepts the postulate of trichotomy and the two postulates of transitivity he may, and often does, accept the further prudential principle of maximizing the *so ordered* preferences. It is, however, possible that some people accept weaker postulates for ordering their preferences and hence different principles of maximization. If so, we should be confronted with divergent principles of prudence at an earlier and more primitive stage of prudential deliberations than is normally assumed. In this connection it is relevant to draw attention to a remark by W. Edwards that the postulates of trichotomy and transitivity on the one hand, and the maximizing principle on the other might come into conflict, since "it might become too costly an effort . . . to maintain a weakly ordered preference field." (See Edwards [1].)

---

[3] Kant himself expresses the two sides of the interdependence by his doctrine of the primacy of practical over pure reason and by his thesis that "ought" implies "can."

Starting with the acceptance of prudential maximization on the basis of the orthodox postulates, one may extend our set of prudential principles in a variety of directions. In each of them one will soon be faced with the need to choose between prudential principles which are incongruous with each other. For example, if we assume that every member of a community has his own prudential preference ordering and if on this basis we wish to maximize the welfare of the community, we are faced not only with the choice between different welfare functions but also with at least one unsuspected, deepseated, and far-reaching antinomy, which makes the choice a difficult one. Similarly, if after the fashion of F. P. Ramsay and of von Neumann and Morgenstern we introduce an interval scale (corresponding to the Fahrenheit or Celsius scale of temperature) for the measuring of utility, the problem of maximizing it can again be solved by accepting a variety of different principles of maximization. (For a survey of the literature see the bibliography of Edwards [1] and [2].) There is no need to go here into details about the variety of prudential principles which have been actually accepted by different people or which have been proposed for acceptance by theoreticians. Nor do we have to raise, much less solve, the problem of deciding in every case whether a person's decision violates a set of strict prudential principles or conforms to a set of rather lax ones.

9. ON THE COHERENCE OF PRACTICAL ATTITUDES AND THEIR ULTIMATE EVALUATION

In emphasizing that prudential principles apply only to practical attitudes which are, or are treated as if they were, morally indifferent, the notion of moral indifference was left unanalyzed. However, a more detailed examination of practical coherence-principles, especially of their rôle in a person's ultimate or unqualified evaluation of his practical attitudes, constitutes a first step toward the analysis of moral coherence and, hence, of moral evaluation. Let us, then, consider a practically coherent system; that is to say, assume that the practical attitudes of a person conform to a set of principles (expressing secondary reflexive attitudes), which in the sense explained above (§7), are comprehensive, selective, not strongly discordant with (not negatively dominated by) any attitude of higher level, and congruous with each other. Among the questions which naturally arise about the structure of such a system, the following are particularly relevant to our task. First, does such a system contain undominated—i.e., not merely negatively undominated—attitudes? If so, secondly, do some of these undominated attitudes dominate all practical attitudes? And if so, thirdly, how does the domination of all practical attitudes by some undominated attitudes manifest itself?

The answer to the first question is that there must be undominated attitudes if there is to be no unlimited progression of reflexive attitudes of ever higher level. The second question is empirical. However, if a person's practical attitudes, his reflexive attitudes directed to them and his higher reflexive attitudes form a practically coherent system, there must be in it attitudes of second or higher level which, either directly or *via* lower reflexive attitudes, establish this systematic connection. Indeed, if in a practically coherent system a comprehensive secondary attitude is not negatively dominated, it is either not dominated at all, or else positively or indifferently dominated by an undominated higher attitude which, *via* the dominated secondary attitude, dominates all practical attitudes of the system.

This brings us to our third question, namely the manner in which an undominated reflexive attitude can dominate all the practical attitudes of a person. To concentrate on the essential parts of the question it will be useful to introduce some special assumptions which will simplify the discussion without any real loss of generality. It will be convenient to assume that the undominated reflexive attitudes are secondary, since what will be said about them is easily extended to undominated attitudes of third or higher level which dominate practical attitudes only indirectly *via* lower level attitudes.

It will further be convenient to consider only special cases of preference and indifference, namely the absolute pro-attitudes ($A+p=Ap>\neg p$), anti-attitudes ($A-p=Ap<\neg p$), and attitudes of indifference ($A\pm p=Ap\sim\neg p$), since these cases are of special relevance to ethics and since the introduction of the more general, preferential structure is merely a matter of technical refinement. Lastly, I shall not—apart from mentioning it here—consider the possibility that one and only one secondary attitude dominates all of a person's practical attitudes. This is the case of somebody so completely self-satisfied, self-dissatisfied, or indifferent that he has only one (pro-, anti-, or indifferent) attitude toward all his practical attitudes.

Everybody else's—I suspect everybody's—primary practical attitudes fall into three mutually exclusive and jointly exhaustive classes, according to whether he has an undominated pro-, anti-, or indifferent attitude toward them. A practical attitude which is the object of an undominated, reflexive attitude will be called "ultimately evaluated" and, according to whether the reflexive attitude is a pro-, an anti-, or an indifferent attitude, "ultimately approved," "ultimately disapproved," or "ultimately (judged) indifferent." The class of ultimately evaluated attitudes has thus the three mutually exclusive subclasses of ultimately approved, ultimately dissapproved, and ultimately indifferent attitudes.[4]

Before examining the constitution of these classes it will be useful to extend our symbolic abbreviations in the following respects. We add the subscript '2' to the symbols '+', '−', '±' for a secondary attitude to indicate that it is undominated; and we use '*' ambiguously for '+', '−', '±'. Thus $A+_2(A-p)$ means that A has an undominated pro-attitude toward his anti-attitude which is directed toward p and $A*_2(A\pm p)$ means that A has an undominated secondary attitude toward his indifferent attitude which is directed toward p. Lastly, whenever it is important to distinguish explicitly between propositional variables and propositional constants, the former will be indicated by small Roman letters without subscripts, e.g., 'p', the latter by such letters with subscripts, e.g., 'p_0'.

A person's ultimate evaluation of one of his practical attitudes toward a situation or event, schematically $A*_2(A*p_0)$, could in principle be either an evaluation by criteria or a concrete evaluation. In the former case $A*p_0$ is evaluated by A as a practical attitude directed toward a kind of situation or event, i.e., toward p_0 as satisfying a certain necessary and sufficient or merely necessary condition or criterion. The criterion may be a simple or conjunctive attribute or a disjunction of such attributes. If the criterion is a necessary and sufficient condition, the ultimate evaluation can be schematically represented as follows: For all p—$A*_2(A*p)$ logically implies, and is logically implied by $p\epsilon C_1$ or . . . $p\epsilon C_n$ (where "$p\epsilon C$" stands for "p is an instance of C"). It may, for example, be that A has an undominated like for his dislike of any situation or event which involves an exhibition on his part of pride, envy, avarice, lust, gluttony, anger, or sloth. If the criterion is merely a necessary condition the logical implication is unilateral.

A concrete evaluation is one that defies the employment of criteria, because what is evaluated is a practical attitude toward a particular situation or event in all its concreteness. It should be noted that strictly speaking the schema $A*_2(A*p_0)$ does not adequately represent a concrete evaluation since no proposition p_0 is capable of fully describing a concrete situation or event with its unlimited properties and relational characteristics. Whether or not a person's ultimate evaluations are or are not concrete is an empirical question. There is, of course, no doubt that the members of any society employ evaluations by criteria in order to facilitate cooperation and to make public (for example, judicial) checks available and controllable. What may be doubtful is whether these evaluations are ultimate or whether, like prudential evaluations, they are subject to restricted or qualified principles of practical coherence only.

It seems that most people in our culture employ some negative and some very general positive criteria of ultimate evaluation. Most of them would ultimately disapprove of their liking for any action *qua* cruel action. And there are, I believe, very few people who do not have an undominated pro-attitude toward their liking of any action in so far as they regard it as belonging to the class of actions which decrease unnecessary suffering or which increase our knowledge of man's place in the universe. However that may be, only principles of ultimate evaluation by criteria, that is to say principles to the effect that for all p—$A*_2(A*p)$, impose, alone or together with similar principles, a condition on all practical attitudes and are, therefore, principles of practical coherence in the sense of our definition.

---

[4] The ambiguity of the term "indifferent" is always resolved by the context in which it occurs.

10. On the Ultimate, Transpersonal Evaluation of Practical Attitudes and Their Moral Coherence

So far we have considered a person's ultimate evaluations of his own practical attitudes only. However, his undominated reflexive attitudes do not lose their reflexive character by being directed to the practical attitudes of a wider class of persons, provided that he himself belongs to it. Schematically, $A^*_2(X^*p)$ is no less reflexive than $A^*_2(A^*p)$, provided that the personal variable X ranges over a class of persons of whom A is one. A's ultimate "transpersonal" evaluation is made from the social point of view, if X ranges over all members of A's community, and from the universal human point of view if X ranges over all human beings. To be more specific $A+_2(X^*p)$, $A-_2(X^*p)$, $A\pm_2(X^*p)$ respectively express that from the social (or universal human) point of view A ultimately approves of X^*p, disapproves of it, or judges it to be indifferent.

It seems to me a rather unimportant verbal question, whether a person's transpersonal, ultimate evaluations of practical attitudes should be called his "moral evaluations" only if X ranges over all human beings, or also if it ranges merely over the members of his community. One might, on the one hand, feel inclined to call a person's ultimate evaluations from the social point of view "moral" if they are his only ultimate evaluations. There seems little reason for denying the name of a morality to the undominated reflexive attitudes of a Greek living in the 5th century B.C., whose personal variable X did not range over the barbarians. On the other hand, by requiring X to range over all human beings, one is, I hope, in better agreement with modern usage. By adopting it, as I shall do henceforth, one moreover emphasizes that morality is *ipso facto* social since if X ranges over all human beings, it ranges over the members of any human community.

A person's moral attitudes are, thus, his undominated, reflexive attitudes directed toward everybody's practical attitudes including his own. By these reflexive attitudes, provided that they are congruous with each other, his and everybody else's practical attitudes are evaluated as moral $(A+_2(X^*p))$, as immoral $(A-_2(X^*p))$ or as morally indifferent $(A\pm_2(X^*p))$. If these reflexive attitudes are comprehensive, i.e., if they alone or in conjunction with a finite number of other such attitudes impose a condition on all practical attitudes, they are principles of moral coherence. The evaluation of practical attitudes by (comprehensive) principles of coherence is, as we have seen, evaluation by criteria. Schematically, a principle of moral coherence can, therefore, be represented by a proposition to the effect that for all p and for all X—$A^*_2(X^*p)$ logically implies, and is logically implied by, $p\epsilon C$, or . . . $p\epsilon C_n$. If the criteria for the morality, immorality, or moral indifference of practical attitudes are merely necessary, the logical implication is unilateral.

To stress an important analogy between principles of practical and principles of factual coherence, the former might also be represented in a somewhat different schematic form. Let us write M^{+2}, M^{-2}, $M^{\pm 2}$ respectively for the class of practical attitudes which a person A evaluates as moral, immoral, or morally indifferent; and M^{*2} ambiguously for any of these three classes. Let us further assume that C is a necessary criterion for the practical attitude's X^*p belonging to M^{*2}. The principle of practical coherence which corresponds to this assumption can then be expressed by: "Being an M^{*2}" logically implies "Being a C." And this principle has the same form as a constitutive principle associated with a maximal kind of factual thinking. In particular, M^{*2}, the class of practical attitudes evaluated as moral (immoral, morally indifferent) by a person A corresponds to a maximal kind employed in A's factual thinking; and C, his criterion of morality (immorality or indifference), corresponds to a constitutive attribute of this maximal kind. For example, just as, say, the maximal, factual kind of Whiteheadian events is associated with the constitutive factual principle to the effect that "Being an event" logically implies "Being causally connected with other events," so the maximal kind of moral (morally approved) attitudes as conceived, say, by a utilitarian is associated with the constitutive moral principle to the effect that "Being a moral attitude" logically implies "expressing the intention to increase the happiness of at least one person without decreasing the happiness of anybody else."

Indeed we may, analogously to the definition of a categorial framework of factual thinking, define the notion of a moral framework (a categorial framework of moral evaluation) as a categorization of all practical attitudes into the maximal kinds of moral, immoral, and morally indifferent ones together with their associated constitutive and individuating principles. As regards the latter principles it seems clear that the individuation of a practical attitude—moral or otherwise—will depend on the individuation of the person having the attitude and thus on the framework of factual thinking which is presupposed by the moral framework. This dependence is only one aspect of the general dependence of practical coherence on factual coherence, which has been pointed out earlier (§7).

It is not possible here to examine the great variety of actual and of some conceivable moral frameworks and of the principles of moral coherence which correspond to their acceptance and employment. Two remarks, however, may serve to indicate this variety. First, a constitutive attribute of a person's class of moral (immoral, morally indifferent) practical attitudes may be either a characteristic which does not involve attitudes or one which involves them. Examples are on the one hand "decreasing the number of starving people," on the other "conforming to the standards of prudence which one would observe in dealing with one's own interests." Secondly, insofar as there is a multiplicity of factual frameworks (in which the situations or events are formulated, toward which the morally evaluated attitudes are directed) there must be a consequential multiplicity of moral frameworks (which contain the constitutive attributes instantiated by these situations or events). This multiplicity must be distinguished from the multiplicity which arises from different categorizations of practical attitudes toward situations or events described in the same factual framework. It is conceivable and, I think, frequently the case that two people sharing the same categorial framework of their factual beliefs employ very different moral frameworks.

11. Moral Coherence and the Concept of Morality

A person's principles of moral coherence, as we have defined them, are principles of practical coherence. They are directed toward all his practical attitudes by imposing, alone or in conjunction with other such principles, conditions on all of them. Their specific content consists, in particular, in the way in which they divide all practical attitudes $X^{*}p$ into three maximal classes M^{+2}, M^{-2}, and $M^{\pm 2}$ of moral, immoral, and morally indifferent practical attitudes by means of necessary and sufficient conditions for membership of these classes, or, at least, by providing necessary conditions for membership of them. Their manner of acceptance consists in the undominated preference for these principles over any others with which they are incongruous. We have arrived at this definition by examining—and asking some natural questions about—stratified systems of attitudes which satisfy the requirements of practical compatibility of all practical attitudes and of congruity of all undominated practical attitudes, while admitting discordances between attitudes of different level. However, there still remains the question whether the definition is empty. This is an empirical question, to which the answer is negative—as I believe on the grounds of introspection and observation (and as I believe experiments of the kind made by decision-theorists would show). Granting all this, we may ask how far the proposed definition of "moral coherence" corresponds both to ordinary and traditional philosophical ethical terminology.

I shall here content myself with briefly indicating that it corresponds fairly closely to the notion of a morality as characterized by a contemporary moral philosopher whose competence, caution, and acuteness is generally respected by the philosophical community. W. K. Frankena (see Frankena [1]) distinguishes between a formal and a narrower, material concept of morality. A person's "action-guide" is moral in the former sense if he (A) takes it "as prescriptive," if he (B) "universalizes it" and if he (C) regards it as "definitive, final, overriding or supremely authoritative." According to the narrower concept of morality a person has a morality "only if perhaps in addition to such formal criteria as (A) and (B)" his action-guide (D) "includes or consists of judgments (rules, principles, ideals, etc.) that pronounce actions to be right, wrong, good, bad, etc., simply because of the effects they have on the feelings, interests, etc., of *other* persons where 'other' may mean 'some other' or 'all other' and where the class of persons included may include the person who is using the action-guide."

The definition of moral coherence satisfies these criteria. Their acceptance is, moreover, justified by the analysis of practical and reflexive attitudes which led up to the definition: (A) The reflexive attitudes expressed by principles of moral coherence are "prescriptive" in being directed to practical attitudes with which they may be in concordance or discordance. They are moreover self-prescribed or autonomous in a sense which is characteristic of moral as opposed, for example, to legal prescriptiveness. I should even be prepared to argue that the discordance of an undominated reflexive with a primary practical attitude goes a long way toward accounting for the feeling of conflict between duty and inclination. (B) needs no comment. (C) The definitiveness, finality, and overriding or supremely authoritative character of moral "action-guides" is expressed and explained by the undominated domination of a person's principles of moral coherence over all his practical attitudes. (D) The social content of principles of moral coherence follows, as has been pointed out earlier, from the fact that if *A*'s personal variable *X* ranges over all human beings, it ranges over the members of *A*'s community. Insofar as man is a social and political animal, anybody's reflexive concern with himself as an exchangeable representative of mankind cannot fail to have social content.

The preceding account of moral coherence might possibly provoke the vague charge of an unwarranted subjectivism or relativism which makes morality a "mere" matter of attitudes. Since this charge cannot mean that moral attitudes should not be regarded as attitudes, it probably implies that they are not mere matters of taste as are many primary likes and dislikes. However, on the preceding analysis, which in taking the stratification of attitudes seriously differs fundamentally from the so-called "attitude-theories" of ethics, moral attitudes are *not* primary attitudes at all, but undominated, reflexive attitudes of a special kind. A possibly instructive way of answering the charge of unwarranted subjectivism and similar charges is to compare principles of factual coherence such as the principle to the effect that "being an event" logically implies "being causally connected with other events" with principles of moral coherence such as the principle to the effect that "being a moral (morally approved) attitude" logically implies "expressing the intention to increase the happiness of at least one person without decreasing the happiness of anybody else."

The most obvious differences between principles of factual and principles of moral coherence are that whereas the former apply to all factual beliefs of a person, the latter apply to all his practical attitudes; and that whereas the former express comprehensive factual beliefs, the latter express comprehensive reflexive attitudes. By attending only to these features one might get the wrong impression that the principles of factual coherence are uncontaminated by attitudes, and that the principles of moral coherence are free from factual content. A retrospective glance at the preceding analyses must dispel this impression. They show quite clearly that *both* kinds of coherence-principle presuppose the concepts of empirical and hence of logical compatibility. In particular, the notion of practical compatibility is reducible to that of empirical and hence of logical compatibility; and the notion of congruity of reflexive attitudes is partly defined in terms of practical compatibility (and partly in terms of concordance).

The preceding analyses show, equally, that the acceptance of either kind of coherence-principle is a matter of attitude in that both are accepted in preference to alternatives. Principles of factual coherence are accepted in preference to principles with which they are logically incompatible—the preference being primary and (positively or indifferently) dominated by principles of practical coherence. Principles of moral coherence are accepted in preference to principles with which they are incongruous—the preference being reflexive and undominated. Thus both kinds of coherence-principle involve factual beliefs *and* attitudes. To contrast them as if they exemplified an allegedly fundamental opposition of beliefs without attitudes and of attitudes without beliefs is to block the way toward an adequate understanding of their structure and function.

12. From Coherence to an Absolute Factual Reality and an Absolute Moral Ideal

Our examination of the notions of factual and practical coherence has, among other things, exhibited three ways in which systems of factual beliefs and practical attitudes can differ from each other. First, there can be a variety of systems all of which are internally consistent in the precise sense of satisfying the demands of logical, empirical, and practical compatibility, as well as of congruity. Secondly, there can be a variety of systems all of which are internally consistent and which nevertheless diverge from each other in satisfying different principles of factual coherence, of moral coherence, or both. Thirdly, there can be a variety of systems all of which are internally consistent and conform to the same principles of coherence and which nevertheless diverge from each other in their more specific content. Before considering some of the reasons why the second and the third thesis is sometimes denied by philosophers and before considering the relevance of all three of them to the problem of truth, it will be useful briefly to summarize our arguments in their support.

Essentially they were all arguments *ab esse ad posse*. The first thesis seems to be generally accepted and follows in any case from the second. To demonstrate the second thesis—i.e., the variety of coherent systems—for factual beliefs, it was shown that the principles which characterize the categorial framework of a person's factual beliefs are principles of factual coherence and that different such categorial frameworks have been, and are being, employed by different groups of people. To demonstrate the third thesis—i.e., the variety of specific systems, conforming to the same principles of coherence—for factual beliefs, the example of alternative set-theories was considered, whose postulates, though not comprehensive, are highly general. But, of course, more specific differences in factual belief would equally do for our present purpose. Since principles of moral coherence (and of practical coherence, generally) presuppose principles of factual coherence—a person's practical attitudes involving his belief in their factual realizability—the demonstration of the second and third thesis for factual beliefs is *ipso facto* a demonstration of these theses for moral and, in general, for practical attitudes. However, it has been pointed out that in addition two persons sharing the same framework of factual thinking sometimes employ different principles of moral coherence; and that two persons employing the same principles of moral coherence sometimes differ in some of their specific moral attitudes—for example when these attitudes require "prudent" conduct.

Contrary to the second thesis, and in spite of the historical and anthropological evidence in its favor, some philosophers tend to argue that the very nature or possibility of a person's having factual beliefs, moral attitudes, or both, necessitates one, and only one, set of principles of factual coherence, of moral coherence, or of both. One of the reasons for such arguments is a confused notion of coherence, in particular a notion of intuitive or total coherence. Another is what might be called the "transcendental illusion." It consists in mistaking one's ignorance of alternatives for their impossibility; in a *non-sequitur* which overlooks the need for a uniqueness-proof where one is needed; in a *petitio principii* which defines the nature of factual beliefs or practical attitudes so narrowly as to imply an allegedly unique set of coherence-principles; in the use of a special non-logical inference whose nature is, and is left, obscure; or in a mixture of all these features.

Contrary to the third thesis, some philosophers go so far as to argue that the very nature or possibility of a person's having factual beliefs, moral attitudes, or both, necessitates one, and only one, complete set of factual beliefs, moral attitudes, or both—the set being complete in the sense that the addition of any factual belief or practical attitude, not already contained in the set, would render it internally inconsistent. The main reason for such arguments might be called the "dialectical illusion." It consists in the assumption that there is—in fact or in principle—available a special informal or dialectical Logic which reveals the alleged necessary connections. However, from the invalidity of transcendental arguments to the effect that the very possibility of a person's having factual beliefs, moral attitudes, or both, necessitates a unique set of principles of factual coherence, of moral coherence, or of both, there follows *a fortiori* the invalidity of dialectical arguments to the effect that the very possibility of a person's having factual beliefs, moral attitudes, or both necessitates one, and only one, complete set of factual beliefs, practical attitudes, or both.

Is there, then, no absolute (mind-independent) factual reality and no absolute moral ideal? Our analyses imply nothing of the sort. They are, in particular, compatible with two mutually exclusive metaphysical positions. One is a correspondence theory according to which one, and only one, set of factual beliefs corresponds to an absolute factual reality and one, and only one, set of practical attitudes to an absolute moral ideal. The other is a perspectivist theory according to which every coherent system of factual beliefs represents a perspective of an absolute factual reality and every coherent system of practical attitudes a perspective of an absolute moral ideal. Of these two thought-possibilities each will appeal to different people. The second will, for example, appeal mainly to those who, while recognizing the multiplicity of mutually exclusive coherent systems and principles of coherence and while enjoying their contemplation, nevertheless hanker after absolute truth and absolute values. Yet, although the analysis has left room for a choice, the choice itself has to be made on other grounds.

The University of Bristol AND
Yale University

Received January 12, 1971

REFERENCES

[1] EDWARDS, Ward "The Theory of Decision Making," *Psychological Bulletin*, vol. 51 (1954).

[2] EDWARDS, Ward "Behavioral Decision Theory," *Annual Review of Psychology*, vol. 12 (1961) (both [1] and [2] are reprinted in *Decision Making*, selected readings ed. by W. Edwards and A. Tversky) (London, 1967).

[1] FRANKENA, W. K. "The Concept of Morality," *The Journal of Philosophy*, vol. 63 (1966), pp. 688–696.

[1] KÖRNER, Stephan *What is Philosophy?—One Philosopher's Answer* (London, 1969).

[2] KÖRNER, Stephan *Categorial Frameworks* (Oxford, 1970).

[1] LUKASIEWICZ, Jan "On Determinism" (English version of a rectoral address delivered in 1922 at Warsaw University), *Polish Logic 1920–1939* ed. by S. McCall (Oxford, 1967).

American Philosophical Quarterly
Volume 9, Number 1, January 1972

II. DESCARTES ON HIS SUBSTANCE AND HIS ESSENCE

DAVID S. SCARROW

IN his "Synopsis of the Six Following Meditations" Descartes writes:

> . . . we must conclude from all this that things which we clearly and distinctly conceive to be diverse substances, as we conceive the mind and the body, are in fact substances which are really distinct from each other. . . . (HR I, 140–141.)[1]

To conceive of "mind and body" as being diverse substances is to conceive of *a* disembodied mind, i.e., a being which has mental properties but no corporeal properties, and *a* mindless body, i.e., a being which has corporeal properties but no mental properties. To say that the mind and the body are conceived "clearly and distinctly" to be diverse substances is at the very least to say that there is nothing self-contradictory in the concept of a disembodied mind and in the concept of a mindless body. But to say that the mind and the body "are in fact substances which are really distinct from each other" is to claim

(1) that there is a disembodied mind and a mindless body,
(2) that every mind is a disembodied mind and every body is a mindless body.

How can Descartes pass from the premiss that disembodied minds and mindless bodies are logically possible to the conclusion that they are actual and in fact the only kinds of minds and bodies? This passage from possibility to actuality has puzzled many of Descartes' critics. My aim in this paper is to show how Descartes' understanding of the term "substance" and related terms made this passage seem to him perfectly justifiable and made it necessary to conclude that the mind and the body "are in fact substances which are really distinct from one another."

I

Often when Descartes discusses substance and substances he is presenting a version of Aristotelian ontology. Substances are things like Socrates and Plato. Modes are things like Socrates' wisdom or Socrates' justice. Although substances and modes are logically—and hence ontologically—distinct, what it is to be a substance and what it is to be a mode can be understood only by contrasting one with the other. One way of making the contrast is in terms of independent existence. A substance exists without logically requiring the existence of anything else (except God); a mode, on the other hand, requires the existence of something, a subject in which it exists. Thus Descartes says:

> . . . whatever by the power of God . . . is capable of existing without a subject, must . . . be termed substance. (*Reply to Objection VI*, HR II, 250.)

and

> . . . the nature of a mode consists in this, that it can by no means be comprehended except it involve in its own concept the concept of the thing of which it is a mode. . . . (*Notes on a Program*, HR I, 440.)

The contrast between substance and mode can also be drawn in terms of the contrast of complete and incomplete beings. Substances are complete beings which can be conceived in themselves; whereas modes are incomplete beings which must be conceived of through something else.

> . . . neither can I understand justice apart from a just being, or compassion apart from the compassionate; nor may I imagine that the same being as is just cannot be compassionate. But yet I understand in a complete manner what body is, that is to say, I conceive of body as a complete thing. . . . (*Reply to Objection I*, HR II, 22.)

In making these contrasts between substance and mode Descartes took two important positions on ontological issues. First, he rejected both platonism and nominalism. Modes such as Socrates' wisdom or Socrates' justice are not eternal, separable Forms. They are not substances. On the other hand, they are not nothing as nominalistic doctrine implies. They are real, but they are dependent and incomplete. They have the same sort of reality as the surfaces of a cube. Those surfaces exist; in fact there are six of them. But they have a dependent existence. The surfaces are modes of the cube.

---

[1] All references are to *Philosophical Works of Descartes*, ed. by E. S. Haldane and G. R. T. Ross (two vols., Cambridge, 1931).

A second position Descartes took is: If anything exists, substances exist. This, of course, had long been a commonplace in philosophy. The idea that a property might occur at a particular space-time region without being the property of some subject —something which could have other properties as well and which could change its properties over time—did not occur to Descartes or to anyone else in the seventeenth century. Indeed, the doctrine that substances are the presuppositions of all existence had solid Aristotelian roots and formed one of the parameters of seventeenth century philosophical disputation. No one was prepared to question the presupposition that there are substances. Accordingly Descartes was able to use this acceptance as he developed the concept of substance in his own way.

Another philosophical commonplace which Descartes traded upon was the ancient and honorable doctrine that every substance has an essence. Descartes used a variety of terms to speak of a substance's essence. As a rule "essence," "principal property," "principal attribute," and "nature" are interchangeable terms in Descartes' philosophical writings. But however Descartes varied his terminology, he repeatedly insisted upon the traditional doctrine that each substance has one and only one essence.

> . . . there is always one principal property of substance which constitutes its nature and essence, and on which the others depend. (*Principles*, Part I, LIII, HR I, 240.)

> attributes . . . of which neither is contained in the concept of the other . . . [cannot be] coexistent in one and the same subject, for that is equivalent to saying that one and the same subject has two different natures, and this involves a contradiction. (*Notes on a Program*, HR I, 436.)

Descartes followed the tradition of treating essence as logically different from modes. Whereas Fido's whiteness is a distinct though dependent entity and Fido's being white implies the existence of such an entity, Fido's being a dog does not imply the existence of an essence, Fido's dogness, which is distinct from Fido and yet dependent upon him. Being a dog is being a substance with a certain nature or essence. But the only reality is the individual substance. There is no distinct entity, dogness, which Fido "has." This clearly is the view which Descartes has in mind when he writes:

> Thought and extension . . . must not be considered otherwise than as the very substances that think and are extended, i.e., as mind and body. . . . (*Principles*, I, LXIII, HR I, 245–246.)

and when he says that these principal attributes are not "inherent in the substances, as in subjects diverse from themselves" (*Notes on a Program*, HR I, 435).

But although they are distinct, modes and essences are logically related. The modes which a substance can have are determined by the kind of substance it is. A man can be just; but an apple cannot be just. Or, as Descartes would say, a thing whose essence is extension can admit of diverse modes; but it cannot admit of every possible mode.

> For example, it cannot admit of modes of thought such as affirmation and negation. For modes of different attributes cannot coexist "in one and the same subject" (cf. *Notes on a Program*, HR I, 435).

II

These Cartesian dicta on substance, essence, and mode express views which can be found in Aristotle, and which even today can be taken as fundamental to an adequate philosophical logic or philosophical ontology.[2] But Descartes' claims about substance, essence, and mode were not consistently Aristotelian. They often involved a subtle shift in the traditional doctrines, a shift which yielded a radically new philosophical ontology. In this section I shall note three such departures from the traditional doctrines. In the next section I shall show the role of these departures in Descartes' arguments for dualism. Finally, I shall attempt to show what lay behind Descartes' departures from the traditional doctrine of substance, essence, and mode.

(1) One such departure is evident from the very beginning of the discussion of substance in the *Principles*

> By substance, we can understand nothing else than a thing which so exists that it needs no other thing in order to exist. (*Principles*, I, LI, HR I, 239.)

At first this statement may seem to be just a way of defining substance by contrasting it with a mode.

---

[2] Cf. E. Anscombe and P. T. Geach, *Three Philosophers* (Ithica, 1961), ch. I, "Aristotle: The Search for Substance."

A mode, e.g., Socrates' wisdom, does need another thing to exist, namely, Socrates. So the mode is not a substance, but Socrates is a substance. Sometimes, as we have seen, Descartes does say something like this. But this is not what he is saying here. Here he is offering a much more restrictive concept of substance than that which simply contrasts substance with mode. He is not saying simply that an entity is not a substance if it needs a subject in order to exist. He is saying that an entity is not a substance if it needs anything else in order to exist. And this condition disqualifies Socrates as well as Socrates' baldness from being a substance.

Descartes explicitly recognized this consequence of his definition. He recognized that he could not speak of both Socrates and God as substances in the same sense of "substance."

> The word substance does not pertain *univoce* to God and other things. (*Principles*, I, LI, HR I, 239.)

Only God satisfies the original definition of substance. A modified definition must be introduced if we are to speak of Socrates or any man as a substance. It is necessary to distinguish between uncreated substance, God, and created substances.

> Created substances . . . need only the concurrence of God in order to exist. (*Principles*, I, LII, HR I, 240.)

But even this revised definition cannot save the substantiality of Socrates and other ordinary individuals. For whether the "need" in the definition be interpreted in a logical or in a causal sense, ordinary individuals always need something beside God in order to exist.

When "need" is interpreted in the logical sense, this definition says that something is a substance only if it is logically possible that only it and God should exist. But this implies that nothing which has substantial parts can be a substance. For if *X* is a substantial part of *Y*, *Y* cannot exist without *X*. It is not logically possible for only God and *Y* to exist while *X* does not exist. But Socrates' hand is a part of Socrates and, Descartes admits, (*Reply to Objection IV*, HR II, 99), it is a substance. Hence Socrates is not even a created substance. And of course similar arguments could be made for most other ordinary individuals.

When "need" is interpreted in the causal sense, this definition says that something is a substance only if it is causally possible that it should exist alone except for God. This means that nothing that is causally brought into existence by something else in nature will count as a substance, and this consequence is consistent with what Descartes says about substance in some other places, namely, that it cannot be naturally generated or destroyed. (Cf. *Synopsis of the Six Following Meditations.*) Obviously neither Socrates nor most of the other things we might naturally think of as created substances can be counted as such when Descartes' definition is interpreted in this way.

2. Accompanying this tendency to deny substantiality to ordinary individuals such as Socrates is Descartes' repeated suggestion that substance is something which we do not perceive and which is hidden behind the attributes and properties which we do perceive. That these attributes and properties belong to a substance is something which we infer, not something that we perceive.

> . . . when we perceive any attribute, we therefore conclude that some existing thing or substance to which it may be attributed, is necessarily present. (*Principles*, I, LII, HR I, 240.)

We draw this conclusion by appealing to the principal, known by the "natural light" that "no qualities or properties pertain to nothing" (*Principles*, I, XI, HR I, 223); that "a real attribute cannot be an attribute of nothing." (*Arguments Demonstrating the Existence of God*, HR II, 53.) Substance, Descartes here maintains, is something whose existence we "conclude." And in his reply to Hobbes Descartes makes this point again:

> . . . we do not apprehend the substance itself immediately through itself, but by means of the fact that it is the subject of certain activities. . . . (HR II, 64.)

When Descartes argues in this way he is not making an epistemological claim. He is not arguing that we don't perceive substance because we don't perceive anything outside ourselves. On the contrary, in these contexts Descartes allows that we may perceive attributes. His concern is to insist that we do not cognize the subject of these attributes in the same way. Even if we perceive extension, we do not perceive the subject or substance which is extended. The contrast with the traditional account of substance is clear. There is nothing in that account which implies or even suggests that substance is in principle imperceptible.

(3) A third departure from traditional views of substance is found in Descartes' tendency to identify a substance and its essence and to think of the dependence of a mode upon a substance as duplicated in a dependence of mode upon essence. The first tendency—the tendency to identify a

substance and its essence—comes out clearly in the following passage from *Notes on a Program.*

> . . . the extension of a body can within itself admit diverse modes . . . but extension itself, which is subject of those modes is not itself a mode of material substance. (HR I, 435.)

Here Descartes says that extension is "the subject" of modes. But that which is the subject of "any property, quality, or attribute" is a substance (*Arguments Demonstrating the Existence of God,* HR II, 53.)

The second tendency—the tendency to think of modes as being intimately dependent upon essence—appears regularly in Descartes' discussions. For example, in the *Reply* to Mersenne Descartes distinguishes between "unity of nature" and "unity of composition." The former expression applies when two properties belong to the same substance; the latter expression applies when two substances are joined together.

> . . . we clearly perceive that to the same substance to which the possibility of having figure belongs, the possibility of moving also belongs, so that what is figured and mobile is one by the unity of nature. . . . But we do not perceive the same in the case of the thing which we regard under the form of bone, and of that which we view as flesh; hence we cannot take these to be one and the same by unity of nature, but only by unity of composition, viz., in so far as the animal possessing bone and flesh is one and the same. (HR II, 242.)

In this argument Descartes is preparing the way for explaining his position on mind and body. If that preparation is going to be adequate, the argument must be read as assuming the principle that modes are related to one another either as motion and figure are related or as bone and flesh are related. Motion and figure are related by what could be called "positive compatibility," i.e., anything which is modified by being figured could be modified by being in motion. Flesh and bone, on the other hand, are related by "positive incompatibility," i.e., anything which is modified by being flesh cannot also be modified by being bone. What Descartes needs to assume to support his position on mind and body is that any two modes are related by positive compatibility or by positive incompatibility. Given any two modes we can tell by inspection whether they can or cannot modify the same subject.

Descartes' justification for this assumption is an even more basic principle, namely, the principle that the modes of a substance are all related to its essence by implication. Any mode of a substance imples what the essence of that substance is. This is a much stronger assumption than the position mentioned earlier that the essence of a substance determines what kinds of modes it can have. But it is this stronger assumption which is clearly invoked in *Principles,* I, LIII when Descartes says:

> . . . extension in length, breadth, and depth constitutes the nature of corporeal substance . . . all else that may be attributed to body presupposes extension, and is but a mode of this extended thing. . . . (HR I, 240.)

For in saying that whatever is attributed to a body presupposes that it is extended and that extension therefore constitutes the nature, the essence, of corporeal substance Descartes is assuming that each mode of a substance implies what is the essence of that substance.

The same assumption is to be found in *Notes on a Program.*

> . . . the nature of a mode consists in this, that it can by no means be comprehended, except it involves in its own concept the concept of the thing of which it is a mode. . . . (HR I, 440.)

Insofar as Descartes is simply saying that to refer to a particular mode we must refer to the substance whose mode it is (e.g., Socrates' wisdom) his statement expresses a traditional view about the logical properties of modes. But Descartes seems to be saying more than this. He seems to be saying that to understand what a particular mode is—what wisdom is—we must think of the substance whose mode it is. We must do so, it would seem, because the mode implies the substance. But this is a view which is not to be found in the tradition.

Not only did Descartes assume that the modes of a thing imply its essence; he at times suggested that the essence implies or in some way strongly determines the modes. Thus in *Notes on a Program* he says:

> Thought itself, being an inward source from which these modes arise and in which they are inherent . . . [is] an attribute which constitutes the nature of substance. (HR I, 436.)

If we think of an attribute or essence in an Aristotelian way as that which determines what a thing is, we should be puzzled by Descartes' suggestion that the attribute thought gives rise to all the modes of a subject. For this implies that every thinking substance has identical modes—since their nature or essence is identical. But the paradox of this

335083

interpretation is simply another sign that Descartes is not always conceiving of attribute and essence in the traditional way.

III

As the *Objections* to the *Meditations* clearly show, Descartes' arguments for the proposition that there are only two kinds of substances, minds and bodies, and that he, himself, was a purely mental substance perplexed some of his most acute critics. And recent literature shows that contemporary philosophers continue to find the arguments perplexing and challenging. My suggestion is that Descartes' arguments can best be interpreted when they are regarded as assuming the principle of the intimate relation between modes and essences which has just been discussed. More specifically, I suggest that these arguments assume the following three principles:

(I) For each subject *S*, there is a true proposition "*S* is *E*" which gives the essence of *S*.
(II) "*S* is *E*" gives the essence of *S* only if it is implied by each true proposition "*S* is *P*."
(III) "*S* is *E*" gives the essence of *S* only if it implies the falsity of some propositions "*S* is *P*."

The first principle states that each subject has an essence. When this is conjoined with the second principle it implies that the essence of a subject is implied by each of its modes. The third principle eliminates transcendental properties such as *being* and *being self-identical* from being possible essences. For example, even though every true proposition "*S* is *P*" implies "*S* is identical with *S*," *being self-identical* cannot count as the essence of *S* because that property is compatible with *S*'s having any property whatsoever. The essence of a subject delimits the kinds of properties which it can have.

It is plain how these three assumptions can be used to determine what is the essence of a given subject. Consider the argument which I shall call "Argument A." A certain ball has the property of being a sphere, having a volume of a cubic inch, and being in motion. By (I) we know that it has an essence. By (II) we know that its essence is implied by each of these properties. But, so far as we know, all that is implied by each of these properties is *being extended* and the transcendental properties. But (III) eliminates the latter as possible essences. Hence we must conclude that the essence of the ball is to be extended. We can be as certain of this conclusion as we can be that, except for the transcendental properties, this is the only property which is implied by all the properties of the ball.

This pattern of argument is manifest in a passage in the *Principles* where Descartes for the first time in that work stated his position on the nature of substance.

> But although any one attribute is sufficient to give us a knowledge of substance, there is always one principal property of substance which constitutes its nature and essence, and on which all the others depend. Thus extension in length, breadth, and depth constitutes the nature of corporeal substance. For all else that may be attributed to body presupposes extension, and is but a mode of this extended thing; as everything that we find in mind is but so many diverse forms of thinking. (*Principles*, I, LIII, HR I, 240.)

In claiming that "all else that may be attributed to a body presupposes extension" Descartes is assuming that every true proposition about a body implies that it is a body. Why does he assume this? Because he assumes (A) that every body has an essence; (B) that this essence is implied by each of its properties; and (C) that extension in length, breadth, and depth is the only non-transcendental property which is implied by all the properties of a body. So the subject of those properties has extension as its essence. And this is to say that that subject is body.

Although Descartes' arguments in the *Meditations* make the same assumptions, they are not as apparent. In *Meditation* II Descartes points out that the only properties which he can ascribe to himself with certainty are those which imply that he is a thinking being. That he has various corporeal properties and powers is something which he can doubt, but he cannot doubt that he is thinking. And from this premiss Descartes reaches the conclusion—though he does not draw it until *Meditation* VI—that thought is his essence and that he can exist without having corporeal properties and powers.[3]

---

[3] The crucial paragraph in *Meditation* II is that which begins with the question "But what am I?" and ends with the answer "A thing which thinks." The considerations which Descartes used to answer his question permit him only to answer the question: What properties can I know with certainty that I have? If this indeed is his question, then his answer must be taken to mean: What I can so far know of myself with certainty is that I think.

Thus interpreted, Descartes' answer does not state anything about his essence. And although there are some suggestions in this paragraph that he is trying to state something about his essence, particularly the claim that thought "alone cannot be

How does Descartes get from the premiss that he can think of himself without corporeal properties and powers to the conclusion that he need not have corporeal properties and powers? Arnauld objected that the fact that I can think of a right triangle as having a base whose square is unequal to the sum of the squares of its sides does not show that there can be such a triangle.[4] How can the fact that I can think of myself without a body show that I can exist without a body?

If there is any answer to Arnauld's criticism it is in the argument in *Meditation* VI. According to Descartes' own testimony, this is the most authoritative statement in the *Meditations* of his reasons for concluding that thought is his essence. It will be convenient to break the argument into three parts:

> (A) . . . because I know that all things that I apprehend clearly and distinctly can be created by God as I apprehend them, it suffices that I am able to apprehend one thing apart from another clearly and distinctly in order to be certain that the one is different from the other, since they may be made to exist in separation at least by the omnipotence of God. . . .
>
> (B) . . . and, therefore, just because I know certainly that I exist, and that meanwhile I do not remark that any other thing necessarily pertains to my nature or essence, excepting that I am a thinking thing, I rightly conclude that my essence consists solely in the fact that I am a thinking thing. . . .
>
> (C) . . . yet because, on the one side, I have a clear and distinct idea of myself inasmuch as I am only a thinking and unextended thing, and as, on the other, I possess a distinct idea of body, inasmuch as it is only an extended and unthinking thing, it is certain that this I, that is to say, my soul by which I am what I am, is entirely and absolutely distinct from my body and can exist without it. (HR I, 190.)

In (A) Descartes states a general principle which he proceeds to apply in (B) and (C). The principle he states in (A) is:

> If I can apprehend one thing apart from another clearly and distinctly, then they may be made to exist in separation at least by the omnipotence of God. If they can be made to exist in separation at least by the omnipotence of God, then the one is different from another. Hence, if I can apprehend one thing apart from another clearly and distinctly, then the one is different from the other (i.e., the one is not identical with the other).

The key notion in this principle is the notion of apprehending one thing apart from another clearly and distinctly. How is this to be understood? First, how are we to understand the notions of apprehending "one thing" apart from "another" thing? This suggests that one is thinking of an object X under the description P and an object Y under the different description Q. But if Descartes were to conclude from the fact that one can think of X under one description and Y under a different description that X and Y are different, he would be subject to the objection that one and the same thing can be conceived under logically independent descriptions (for example, the Morning Star and the Evening Star). This objection would be rather like Arnauld's objection. Only whereas Arnauld pointed out that Descartes' thinking of two properties as logically independent does not prove that they are logically independent; this objection would be that Descartes' thinking of X and Y under logically independent descriptions does not prove that they are in fact non-identical.

However, Descartes does not conclude that X and Y are different simply from the fact that one can think of X apart from Y. Rather, he draws this conclusion only from the fact that one can "clearly and distinctly" think of X and Y apart from one another. What is the force of "clearly and distinctly" in this context?

My suggestion is that to clearly and distinctly apprehend X and Y apart from one another is to know that there is no possible essence which both X and Y share—no possible Z such that "X is Z" and "Y is Z" are both implied by the assumed properties of X and Y. And this can be known by knowing (A) that all the properties of X imply the possible essence E but not the possible essence E', and that all the properties of Y imply the possible essence E' but not the possible essence E; and (B) that there is no other property which all the properties of X imply (or that there is no other property which all the properties of Y imply). But if we know (A) and (B), then by (I), (II), and (III) we know that E is the essence of X and E' is the essence

---

separated from me," Descartes can be sympathetically interpreted as not making any direct claims about his essence in this place. Certainly the further program of the *Meditations* does not depend upon his having any view of his essence at this place. The question of his essence arises when Descartes develops his metaphysics in *Meditation* VI. (Cf. HR I, 151–152.)

[4] HR II, 83–84.

of Y, and by the "common notion" that things with different essences are not identical we can conclude that X and Y are not identical.

Thus I have suggested that the argument pattern which Descartes adumbrates in *Meditation* VI incorporates principles (I), (II), and (III) and follows the form which is illustrated in Argument A. And this suggestion is confirmed by the actual development of the argument in what I have labeled parts (B) and (C) of Descartes' argument.

In part (B) Descartes proceeds to apply the argument pattern to the particular case of himself.

(a) I know with certainty that I exist.

This premiss is taken from *Meditation* II, and in it Descartes is not only claiming that he is apprehending a particular substance, namely, himself. He is also indicating that he knows something about this substance—namely, that it has knowledge. And in knowing this about himself Descartes can infer,

(b) I am a thinking thing.

Moreover,

(c) I know that there is no other attribute which is implied by those properties which I know that I have.[5]

Hence, by (II),

(d) No other attribute "necessarily pertains to my nature or essence."

But since by (I) I do have an essence and since by (II) my essence is implied by all my properties, and since by (b) and (c) only thought is implied by all my properties, it follows that

(e) "My essence consists solely in the fact that I am a thinking being."

Having concluded that his essence is thought Descartes goes on to conclude in (C) that

(f) This I is entirely and absolutely distinct from body.

For the properties of body do not imply the attribute of thought, and therefore, by (I) and (II), thought cannot be the essence of body. Therefore any body must be a different substance from me.[6]

IV

The argument in *Meditation* VI, as I have interpreted it, strikes a contemporary reader as highly implausible because it turns on the assumption that any property of a substance implies the essence of that substance. In the next section I shall suggest why this assumption could seem so plausible to Descartes. But first I want to consider a rather different kind of argument for the conclusion that my essence is thought. This argument probably will seem much more plausible than the argument of *Meditation* VI.

Postulate A: P is an essential property of X if and only if X cannot exist without being P.
Postulate B: P is the essence of X if and only if P is the only essential property of X.
Premiss 1: It is logically possible for me to exist even though I do not have a body.
Conclusion 1: Hence (by A) having a body is not one of my essential properties.
Premiss 2: It is not logically possible for me to exist without some sort of consciousness—i.e., without thought.
Conclusion 2: Hence (by A) thinking is one of my essential properties.
Premiss 3: I have no other essential property.
Conclusion 3: Hence (by B) thought is my essence.

The difference between this argument and the argument of *Meditation* VI, at least as I have interpreted it, is this: in *Meditation* VI Descartes infers the essence of a substance from any of its properties. In this argument, on the other hand, we infer the essence of a substance from its existence conditions. Thus in this argument to know some property of a substance is not sufficient for knowing its essence; one must also know what is necessary and sufficient for that substance to cease to be. Thus to know that Fido barks does not tell us anything about Fido's essence; but to know that Fido is a dog and that dogs cease to be when they cease to be alive is enough to know that being alive is one of Fido's essential properties.

---

[5] This is to interpret "I do not remark any other thing . . ." as claiming: I know that there is no other thing. . . . But such a claim to positive knowledge is required by the implication in part (A) of the argument that what is to follow will consist of clear and distinct conceptions, i.e., knowledge.

[6] It is interesting to note that shortly after Descartes has drawn the conclusion that he is distinct from his body in *Meditation* VI he twice more appeals to principles (I) and (II). (A) He argues that such powers as the power of motion cannot be attributed to a soul because "in the clear and distinct conception of these there is some sort of extension found to be present, but no intellection at all. . . ." (HR I, 190). (B) He argues that the power to produce sensations cannot belong to me since "it does not presuppose thought." (HR I, 191). In both of these arguments Descartes is assuming that the properties and powers of a single substance all imply the same essence.

There is some reason to interpret Descartes as employing this second argument as well as the argument in *Meditation* VI. Thus in his reply to Arnauld Descartes says that "nothing without which a thing can still exist is comprised in its essence." (Postulate A.) And Descartes certainly thought that it was evident (as indeed it may be) that *he* could not exist without some form of consciousness—though of course his body could exist without *him*. (Premiss 2.) And of course he would have had no difficulty accepting Postulate B as evident. So the question is: Does Descartes use Premiss 1 as a premiss in some of his arguments for the conclusion that his essence is thought?

Some of his critics interpreted Descartes as using this premiss in *Meditation* II, and they pointed out that he was not entitled to do so. That he could doubt that he had a body did not imply that it was logically possible that he did not have a body. (Arnauld makes this point in *Objection* IV.) But Descartes agreed with this point, and he repeatedly insisted that in *Meditation* II he was not making any claims about his essence and that hence he was not employing as a premiss the claim that it is logically possible for him not to have a body. His proof of this point, he said again and again, did not come until *Meditation* VI. And there, as we have seen, he does not employ Premiss 1 as a premiss in his argument. Rather in that place he proves that it is possible for him to exist without a body by proving that his essence is thought, and he proves this by assuming that thought as a mode of him implies his essence.

In his response to Arnauld, however, Descartes may be invoking something like Premiss 1. Arnauld had brought up the case where someone conceived of a triangle and mistakenly thought of it as not having one of its implied properties. Descartes responded by arguing that in conceiving of mind and body one is conceiving, not of a thing and a property, but of two "complete things." (HR II, 99.) In calling our conception of mind "complete" Descartes meant that the conception of a mind as a being which thinks is logically sufficient to specify a class of individual substances and that the possession of no other properties than modes of thought is implied by this concept. And this seems equivalent to saying that it is logically possible for a mind to exist without a body. Although this is not exactly Premiss 1, it is close to it.

But there are surprisingly few passages in the *Replies* where Descartes asserts as a premiss anything like Premiss 1.[7] Perhaps the closest approach to Premiss 1 is the statement in his reply to Gassendi that "from the fact that we sometimes notice [mind and body] apart, it is highly proper to infer that they are diverse." (HR I, 256.) For to "notice" mind and body apart is to know that it is logically possible for mind to exist without body. But in spite of such passages as this, Descartes' *Replies* give the impression that he never clearly distinguished this second, more plausible argument from the argument which occurred first in *Meditation* VI and to which he repeatedly returned in his *Replies*.[8]

V

In the second section of this paper I have noted three ways in which Descartes' view of substance diverged from an Aristotelian account. In the third section I have interpreted some of Descartes' arguments for dualism as depending upon these departures. In this section I offer a suggestion to explain how Descartes could find these departures so natural and so compelling.

---

[7] It is also surprising to note that all of the objections to Descartes' claim that thought is the essence of a mind are directed to *Meditation* II. None of them are directed to the crucial passage in *Meditation* VI.

[8] Malcolm in his "Descartes' Proof That His Essence is Thinking" (*Philosophical Review*, vol. 74, July 1965) suggests that Descartes derived his conclusion from the principle:

> *x* is my essence if it is the case that if I am aware of *x* then (necessarily) I am aware of myself and if I am aware of myself then (necessarily) I am aware of *x*.

There are two main objections to this suggestion. (1) As Malcolm admits, if this is the operative assumption, then Descartes' conclusion about his essence follows from what he says in *Meditation* II. But Descartes says explicitly in many places that he does not give all the considerations required to draw his conclusion until *Meditation* VI. (2) Descartes' arguments are logical and ontological, not epistemological. They appeal to the logical relations between a thing's properties and its essence. There is simply no textual evidence that Descartes used Malcolm's principle. Malcolm apparently thinks that there is no other way to explain how Descartes could arrive at his conclusion than to attribute to him this principle. But I have noted two other ways that would have led Descartes to his conclusion, and for each of these ways there is some textual support.

Anthony Kenny, in his discussion of Descartes' proof that his essence is thought, also rejects Malcolm's interpretation. But Kenny is content to claim that Descartes' conclusion rests on an argument from ignorance: since I don't know that I have a body, having a body is not essential to me. Cf. Anthony Kenny, *Descartes* (New York, 1968), pp. 79–95.

Part of the explanation turns on the dual role which traditional accounts of substance and essence had always played—roles which prior to the 18th century were confusedly bound up with one another in the minds of most philosophers. One role of an account of substance and essence is that of providing a philosophical logic and ontology. As philosophical logic an account of substance and essence will point out the differences between some concepts and others, e.g., the difference between the concept *dog* and the concept *brown* as they function in the two sentences "Fido is a dog" and "Fido is brown." As a philosophical ontology an account of substance and essence may point out differences between dogs and browns, e.g., that the identity of indiscernibles applies to browns but not to dogs. It may also point out the difference between being a dog and being brown—the difference between being an individual of a certain kind and having a property. And at a very high level of generality it may claim that every entity which has properties is itself an entity of a certain kind. This high-level doctrine could be expressed as: every property qualifies a substance with a definite essence.

The other role the concepts of substance and essence traditionally played was a scientific one. According to Aristotle and the Aristotelian tradition, scientific knowledge is the explanation of properties through essence. Thus the fact that Fido has four legs is explained by the fact that the essence of Fido is *dog*; the fact that Socrates is rational is explained by the fact that the essence of Socrates is *man*. The essence *dog* determines and thus explains the property *being four-legged*; the essence *man* determines and thus explains the property *being rational*. But this program for science has two limitations. (1) It admittedly cannot explain all the properties of a thing. Thus, though Fido's having four legs might be explained by his essence, his having white spots cannot be explained in this way and is therefore beyond the perview of science. This property is one of Fido's accidents. (2) Not everything has an essence which can plausibly be taken to determine any of its properties. For example, the essence of a table is to be a table. But what determines that a table is heavy is not its essence but the wood of which it is made. And what determines that a table has four legs is likewise not its essence but the carpenter who made it. There was a certain plausibility in thinking of the essence of a natural object like a dog or a man as determining its properties. But it is not at all plausible to suppose that the essence of an artificial object determines its properties.

The fundamental reason for Descartes' modification of traditional doctrines of substance and essence was his rejection of this view of scientific explanation. As a visionary scientist, Descartes was convinced (1) that every body—not just natural bodies—was subject to scientific explanation; (2) that every property of a body could be scientifically explained and that therefore there were no "accidents"; and (3) that there was basically one kind of principle which would serve to explain the properties of all bodies. Descartes spoke of this explanatory and determinative principle as the essence of bodies; he identified this essence as extension; and he expressed his scientific vision by saying that the essence of all bodies is extension. But in saying this Descartes was not denying that there were various kinds of bodies—dogs, men, garments, etc. (Cf., for example, *Notes on a Program* HR I, 436, where Descartes mentions in passing that "garments are substances.") He is simply saying that these distinctions into kinds are not of genuine scientific interest. To understand the properties of a thing what is important to know is not the kind of thing it is but what determines those properties. That is why it is important to know that the essence of any body is extension.

Descartes used the traditional language of substance and essence to express his scientific vision. But these terms took on a radically different significance in the cartesian context. In this context "material substance" means the stuff of which individual material objects are made. Moreover, this stuff has a certain nature or essence which determines the properties of the bodies which are made of it. Descartes used the term "extension" to characterize this essence. The doctrine that the essence of all bodies is extension thus is the doctrine that there is one fundamental stuff—extension—from which all bodies are made. Body is corporeal substance (stuff), and "extension in length, breadth and depth constitutes the nature of corporeal substance" (*Principles* I, LIII).

Descartes' claim that extension is the essence of all bodies is not—as it may sometimes seem to be—a logical or ontological analysis of individual kinds. Rather it is a scientific hypothesis about the ultimate stuff of the world. This stuff is thought of as consisting of various portions (volumes) spread through space. Just as ice and liquid water are different states of the one stuff, so the different stuffs we empirically find in the world—wax, water, air, fire, etc.—are to be thought of as different states of extension. And just as it is the nature of water to assume different states under different conditions, so it is the nature of extension to assume its manifold states in the world. Descartes' example of the wax in *Meditation* II expressed this theory of things quite clearly. At first the wax is hard, yellow, sweet-smelling. After it has been warmed by the fire it becomes liquid, colorless, and odorless. But it is the same stuff before and after the transformations. What remains the same is extension, the stuff from which it is made.

We are now in a position to understand how Descartes came to offer those idiosyncratic views of substance, essence, and mode which played such an important part in his argument for dualism. Whereas philosophers in the Aristotelian tradition had based their views of scientific explanation on their philosophical logic and philosophical ontology, Descartes formed his philosophical logic and ontology from his scientific vision that the essence of all bodies was extended stuff. In his arguments for dualism the concepts of substance and essence function as logical or ontological categories. But Descartes' understanding of these concepts was in large part determined by his scientific theory.

Consider first his definition of created substance as that which needs nothing except God in order to exist. As has been noted, this definition rules out the substantiality of an individual such as Socrates both because Socrates has parts and because Socrates comes to be and passes away. But substance, on this definition, must be naturally ungenerable, naturally indestructible, and naturally without parts. And this is just what extended stuff, as Descartes thinks of it, is. It is ungenerable, indestructible, and without any natural divisions, even though it can be divided into infinitely many segments which move relative to one another. Descartes takes extended stuff as a paradigm of created substance and forms his definition accordingly.

Next consider Descartes' view that substance is something which is not perceived but is inferred from properties which are perceived. Again this view rules out ordinary perceivable individuals such as Socrates as substances but it fits in nicely with the paradigm of extended stuff as substance. Just as the stuff which is first ice and then water (a) will be known through its manifest states, (b) will not be definable simply in terms of its manifest states, and therefore (c) will be known by inference from its manifest states—so extension will be known by inference based upon the manifest properties in the physical world. To know the basic stuff of the world requires the use of scientific reason. One can no more know the ultimate physical reality simply by using his senses than he can know the true size of the sun simply by looking at it.

Finally, Descartes was convinced that the properties of the physical world were determined by the geometrical and kinetic properties of the basic stuff, extension. His view was not unlike that of the contemporary physicists and chemists who think of the properties of copper, water, brass, and the other elements, compounds, and mixtures as determined by their atomic and sub-atomic structures. Moreover, just as the physicists and chemists infer the structure of a substance from its observed properties, so Descartes thought that the geometrical and kinetic properties of a portion of extension could be inferred from its manifest properties. Thus in the order of being the structure or stuff determines the manifest properties; while in the order of knowing the manifest properties determine the structure or stuff. When this view is stated in the language of essence and modes, we get: In the order of being the essence determines the modes; while in the order of knowing the modes determine the essence. When we interpret the essence of a body as the stuff out of which it is made, we can understand why Descartes could say and be convinced that a mode implies the essence of which it is a mode. By knowing a mode of a body, he thought, one could always infer the stuff of that body by reasoning that only this stuff could determine such a mode. As we have seen, this doctrine that a mode implies the essence of which it is a mode plays a crucial role in Descartes' arguments for the real distinction between body and mind. Now we can see why this doctrine appeared so plausible to Descartes.

My explanation of Descartes' revisions of the traditional doctrines of substance, essence, and mode has been this: As a scientist, Descartes was convinced that the properties of all bodies were determined by the one stuff of which they were all made, extension. He then was guided by this scientific conviction in formulating his philosophical logic and philosophical ontology. Not that Descartes clearly distinguished between scientific and logical or ontological questions. On the contrary, he was no clearer about this distinction than were his predecessors. Nevertheless, it was his peculiar logic and ontology, not his scientific theory, which Descartes used in most of his arguments for dualism.

VI

I have suggested that we understand Descartes' claim that extension is the essence of every body as saying that every body is made out of the one extended stuff. Should we interpret Descartes' claim that thought is the essence of every mind analogously? That is, should we interpret Descartes as saying that every mind is made out of the one mind stuff?

If we adopted this interpretation then we would have to attribute to Descartes the view that there is one universal mind in which we all participate. Some of Descartes' critics charged him with holding this view.[9] But Descartes emphatically rejected this charge, and he insisted on the real difference between individual minds. Nevertheless, Descartes' disclaimers do not settle the matter, and a survey of his characteristic statements of the view that thought is an essence of the mind shows that they are altogether parallel to his statements about extension as the essence of body. Consider the following passages:

> Thought and extension . . . must not be considered otherwise than as the very substances that think and are extended, i.e., as mind and body. . . . (*Principles* LXIII, HR I, 246.)
>
> Thought itself, being an inward source from which these modes arise and in which they are inherent . . . (is) an attribute which constitutes the nature of substance. (*Notes on a Program*, HR I, 436.)
>
> I tried to remove the ambiguity from the word *thought* in the articles 63 and 64 of the first part of my *Principles.* Just as extension, which constitutes the nature of body, differs greatly from the various shapes or modes which it takes on; so thought, or thinking nature, which I consider constitutes the essence of the human mind, is far from being this or that act of thinking. Mind has from itself the power of choosing this or that act of thought, but not that of being a thinking substance, just as a flame has of itself, in its capacity as efficient cause, the power of extending itself on one side of the other, but not that of being an extended thing. By thought, therefore, I understand not a universal comprehending all the modes of thinking, but a particular nature which receives all the modes just as extension is a nature which receives all the shapes.[10]

In the first and third passage we can understand the claims made about extension by thinking of extension as extended stuff. The passages assume that we shall understand the claims about thought analogously. And the second passage likewise is naturally interpreted as analogous remarks about extension would be. But this interpretation requires us to think in terms of thinking stuff and the universal mind.

Insofar as we can make sense of Descartes' claims that thought is the essence of mind we seem obliged to interpret him as committed to a universal thinking stuff. If we follow Descartes and reject this interpretation, then we shall have to conclude that his principle line of reasoning rests on an analogy which he cannot ultimately accept. Having drawn this conclusion, we must again raise the question: What did Descartes mean when he claimed that thought is his essence? The possibilities for deepening our understanding of Descartes that are implicit in the question have not yet been completely brought to light.

Kalamazoo College

Received October 9, 1970

---

[9] Cf. L. J. Beck, *The Metaphysics of Descartes* (Oxford, 1965), p. 116. Beck's references are to ATiii, 403 and ATiii, 513.

[10] Quoted from Anthony Kenny, *Descartes*, *op. cit.*, p. 68. This is from a letter to Arnauld of 1648. (AT v, 221.) Kenny notes that the quotation also appears in Beck, *op. cit.*, p. 117. Kenny's translation is somewhat different from Beck's.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 1, January 1972

III. IDENTITY THROUGH TIME

DOUGLAS ODEGARD

THERE are three views of identity through time which are worth considering. Two have received a fair amount of attention recently; the third has been badly neglected. One familiar alternative maintains that the concept "identical with" resembles the concept "better than" in two ways: first, the statement "*a* is identical with *b*" is elliptical for some statement of the form "*a* is the same *f* as *b*," where '*f*' functions as a variable and can take different values in different contexts; secondly, there is nothing logically wrong with saying "*a* is the same *F* as *b* but not the same *G*," even though *b* was an *F* and a *G* and *a* is an *F* and a *G*, just as there is nothing logically wrong with "*a* is a better typist than *b* but not a better stenographer," even though *a* and *b* are both typists and stenographers. Peter Geach is the leading spokesman for this view. The other familiar alternative rejects the ellipsis thesis, while admitting that "*a* is identical with *b*" is logically equivalent to "There is a value of *f* for which '*a* is the same *f* as *b*' is true," and denies the possible truth of "*a* is the same *F* as *b* but not the same *G*" under the indicated conditions (as well as under others). David Wiggins is a leading spokesman for this view. The third alternative rejects the ellipsis thesis, with the indicated caveat, but does not reject the possible truth of "*a* is the same *F* as *b* but not the same *G*" under the indicated conditions. If one wanted to attach a philosopher's name to this view, then John Locke would be one of the earlier candidates (see Appendix). For convenience sake, let us call the three views, respectively, "Geach's theory," "Wiggins' theory," and "Locke's theory."

Before going any further, it should be emphasized that our only concern here is with identity through time. The constants '*a*' and '*b*' in "*a* is identical with *b*" pick an individual out as it exists at different times—as in, e.g., "Sandra Jones is identical with Sandra Pirie, since Sandra Jones's maiden name was 'Pirie'." Both Geach and Wiggins develop positions which have a broader scope than this. For example, they are concerned with the logic of statements like "Cicero was identical with Tully" and "Miss Ambrose is identical with Mrs. Lazerowitz."

Geach's theory is characterized by remarks such as

> a baby, a youth, an adult, and an old man may be unrecognizably different, although the same name is borne throughout life. But it is meaningless to say without qualification that the baby, the youth, the adult, and the old man are "the same," or "the same thing," and that this is what justifies us in calling them by the same name. . . . "The same" is a fragmentary expression, and has no significance unless we say or mean "the same *X*," where '*X*' represents a general term. . . . What is implied by our use of the same name throughout a period of years is that the baby, the youth, the adult, and the old man are one and the same *man*.[1]

> I maintain it makes no sense to judge whether *x* and *y* are "the same" or whether *x* remains "the same" unless we add or understand some general term—the same *F*.[2]

> It is surprising that Frege should . . . have constantly assumed that "*x* is the same *A* as *y*" does split up into "*x* is an *A* (and *y* is an *A*)" and "*x* is the same . . . as *y*." We have already by implication rejected this analysis.[3]

> I am arguing for the thesis that identity is relative. When one says "*x* is identical with *y*," this, I hold, is an incomplete expression; it is short for "*x* is the same *A* as *y*," where '*A*' represents some count noun understood from the context of utterance—or else, it is just a vague expression of a half-formed thought.[4]

> I deny that any *one* relation . . . is *the* relation expressed by "is identical with." My whole thesis was that "is identical with" expressed now one, now another, relation, according to the context of utterance.[5]

---

[1] P. T. Geach, *Mental Acts* (London, 1957), p. 69.

[2] P. T. Geach, *Reference and Generality* (Ithaca, 1962), p. 39.

[3] *Ibid*, p. 152.

[4] P. T. Geach, "Identity," *Review of Metaphysics*, vol. 22 (1968–69), p. 3.

[5] "A Reply" (to Fred Feldman), *Review of Metaphysics*, vol. 22 (1968–69), p. 556.

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

If we adopt the notation used by Wiggins, Geach's theory can be expressed as follows:

(a) The statement "$a = b$" is elliptical for a statement of the form "$a \overset{=}{\underset{f}{-}} b$."

(b) There is nothing logically wrong with a statement of the form "$(a \overset{=}{\underset{f}{-}} b)\ \&\ (a \overset{\neq}{\underset{g}{-}} b)\ \&\ (ga\ \&\ gb\ \&\ fa\ \&\ fb)$."

The theory has the following corollaries:

(c) $(a = b) \equiv (\exists f)\ (a \overset{=}{\underset{f}{-}} b)$.

(d) $(\exists f)\ (a \overset{=}{\underset{f}{-}} b)$ does not entail $(\forall f)\ [(fa\ \&\ fb) \supset (a \overset{=}{\underset{f}{-}} b)]$.

(e) The conjunction "$(Ga\ \&\ Gb)\ \&\ (a = b)$" does not entail "$(a \overset{=}{\underset{G}{-}} b)$" unless its second conjunct either is elliptical for "$a \overset{=}{\underset{G}{-}} b$" or is elliptical for some other instance of "$a \overset{=}{\underset{f}{-}} b$" such that the value of f demands the entailment; e.g., say 'G' replaces "horse" and the value of f is "stallion."

An advocate of Locke's theory accepts (b) but rejects (a). Thus, he accepts (c), but on its own merits and not as a corollary of (a). He accepts (d) as a corollary of (b). He accepts the part of (e) which rejects the entailment but he does not accept the need for the proviso. Let me expand slightly on this last difference since it provides one respect in which a Lockean is at an advantage.

According to Geach, if the conjunction "$(Ga\ \&\ Gb)\ \&\ (a = b)$" is elliptical for "$(Ga\ \&\ Gb)\ \&\ (a \overset{=}{\underset{G}{-}} b)$," then obviously it entails "$(a \overset{=}{\underset{G}{-}} b)$." And if the conjunction is elliptical for another statement of the form "$(Ga\ \&\ Gb)\ \&\ (a \overset{=}{\underset{f}{-}} b)$" such that the value of f semantically requires the entailment, then again the entailment holds. But if the value of f neither is G nor semantically dictates the entailment, then the entailment does not hold. A Lockean recognizes that "$(Ga\ \&\ Gb)\ \&\ (a = b)$" *may* be elliptical in the indicated way and would agree that, *if* the conjunction is elliptical, then Geach's proviso is relevant. But a Lockean also holds that "$(Ga\ \&\ Gb)\ \&\ (a = b)$" need not be elliptical in the indicated way and in that case he would say that the entailment simply fails to hold. Granted, since a Lockean accepts (c), he must admit that there is a reason for the truth of "$a = b$" and that a statement of the reason will take the form of "$a \overset{=}{\underset{f}{-}} b$." And it might be that the reason in the given case is that "$a \overset{=}{\underset{G}{-}} b$" is true. But this does not mean that everyone who truly (or even knowingly) asserts "$(Ga\ \&\ Gb)\ \&\ (a = b)$" either says or means "$(Ga\ \&\ Gb)\ \&\ (a \overset{=}{\underset{G}{-}} b)$." For example, he might accept the second conjunct on authority, without knowing whether its truth is based on "$a \overset{=}{\underset{G}{-}} b$." And he could at least raise the question of whether this is so, which is something he could not do if "$a = b$" simply *meant* "$a \overset{=}{\underset{G}{-}} b$." Thus, according to Locke's theory, if "$(Ga\ \&\ Gb)\ \&\ (a = b)$" is to yield "$a \overset{=}{\underset{G}{-}} b$," it must be *augmented* by either "$a \overset{=}{\underset{G}{-}} b$" or an instance of "$a \overset{=}{\underset{f}{-}} b$" in which the value of f is semantically tied to G, whereas for Geach the second conjunct must be understood as *meaning* one of these two things.

Thesis (a) in Geach's theory might be regarded as in part an attempt to account for thesis (b). Wiggins thinks that (b) (which he calls the "relativization thesis" and labels R) constitutes Geach's final rationale for the thesis that if someone tells you that $a = b$, you should always ask "the same what as b?" But (b) is not the most fundamental part of Geach's theory if the theory holds that someone must accept (a) in order to be in a position to accept (b). The reasoning might go as follows: "Thesis (b) implies that '$(a \overset{=}{\underset{F}{-}} b)\ \&\ (a \overset{\neq}{\underset{G}{-}} b)\ \&\ (Ga\ \&\ Gb)$' can be consistent and prima facie this might look curious because the conjunction might seem to entail the inconsistent '$(a = b)\ \&\ (a \neq b)$'. Appearances are misleading, however, because '$a \overset{=}{\underset{F}{-}} b$' entails '$a = b$' and '$a \overset{\neq}{\underset{G}{-}} b$' entails '$a \neq b$' only if the entailed statements are either elliptical for the entailing statements or elliptical for statements of the form '$a \overset{=}{\underset{f}{-}} b$' with special values of f. Thus, '$(a = b)$ & $(a \neq b)$' can be just as consistent as '(*a* is better than *b*) & (*a* is not better than *b*)' when the latter is elliptical for '(*a* is a better swimmer than *b*) & (*a* is not a better father than *b*)'."

A Lockean would dismiss the legitimacy of this line of reasoning. He accepts (b) and therefore must deny that "$(a \frac{=}{F} b)$ & $(a \frac{\neq}{G} b)$ & $(Ga$ & $Gb)$" entails something inconsistent. But he does not do this by arguing that "$(a = b)$ & $(a \neq b)$" is either nonsense or possibly consistent. He accepts the latter conjunction as sensible and inconsistent, and denies that it is entailed by "$(a \frac{=}{F} b)$ & $(a \frac{\neq}{G} b)$ & $(Ga$ & Gb.") For, although by (c) "$a \frac{=}{F} b$" entails "$a = b$," "$a \frac{\neq}{G} b$" does not entail "$a \neq b$." That is $(\exists f)$ $(a \frac{\neq}{f} b)$ does not entail "$a \neq b$," which is to say that, although for some value of f a is the same f as b, a and b can still be the same individual, since there can still be some value of f for which they are the same f.

To take an example, let us stipulate a use for "same corpus" such that *a* and *b* are the same corpus only if all their parts are the same. And let us use "same body" in its pedestrian sense such that *a* and *b* can be the same body without having all the same parts. Thus, the body I now have and the body I had as a baby are the same body but not the same corpus, since some of the molecular parts are different and my former body contained an appendix, whereas my present one does not. Now, Geach and a Lockean might agree that "My current body is the same body as my infant body" entails "My current body and my infant body are the same individual." But they would do so for different reasons, since Geach would introduce a proviso specifying certain interpretations of the entailed statement and a Lockean would reject the need for such a proviso. More importantly, Geach would hold that "My current body is not the same corpus as my infant body" entails "My current body is not the same individual as my infant body" given the same type of proviso, whereas a Lockean would hold that there is a *non-elliptical* use of "My current body and my infant body are not the same individual" in which it is *not* entailed by "My current body is not the same corpus as my infant body." For brevity's sake, let '$\frac{=}{C}$' stand for "is the same corpus as" and '$\frac{=}{B}$' stand for "is the same body as."

The Lockean proposal for rescuing thesis (b) from inconsistency works because for it "$a \frac{=}{C} b$" has a number of implications, including "*a* is a corpus," "*b* was a corpus," "$a = b$" and "*a* and *b* have exactly the same parts." Thus, "$a \frac{\neq}{C} b$," which can be read in such a way that it is true if anything presupposed or otherwise implied by "$a \frac{=}{C} b$" is false, is equivalent to the disjunction "$\sim (a = b)$ v $\sim$ (*a* is a corpus) v $\sim$ (*b* was a corpus) v $\sim$ (*a* and *b* have exactly the same parts) v . . ." and is true if any one disjunct is true. This means that, in the given case, a Lockean can assert "$a \frac{\neq}{C} b$" on the ground that *a* and *b* do not have exactly the same parts, without having to sacrifice "$a = b$."

One way of expressing the critical point here is to say that whereas the first conjunct in "$(a \frac{=}{B} b)$ & $(a \frac{\neq}{C} b)$" is an implicit identity statement, the second conjunct is not an implicit "anti-identity statement." The logic of the first conjunct is not to be assimilated to that of "*a* is the same size as *b*" or "*a* is the same official as *b*" since neither implies "$a = b$." The logic of the second conjunct, on the other hand, is not to be assimilated to that of "This chair is numerically completely distinct from the chair Plato sat on when he wrote the first book of the *Republic*" if the latter is intended to imply that this chair and Plato's chair are not simply different chairs but different individuals—i.e., that there is no concept 'f' under which we can say they are the same f (e.g., "collection of atoms," "piece of matter"). The fact that a Lockean can utilize non-relativized concepts of identity and difference to make points of this sort gives him another advantage over Geach.

Geach does suggest a number of arguments on behalf of thesis (a) and presently I want to show that they are ineffective. First, however, note that if there is nothing which particularly recommends the adoption of thesis (a), it should be abandoned. In support of this, consider three further disadvantages of the thesis. Henceforth, call thesis (a) "relativism," bearing in mind that relativism in this sense is independent of thesis (b) and is therefore distinct from what Wiggins calls "the relativization thesis."

(i) Relativism cannot flatly assert $(a = b) \equiv (\exists f)\ (a \overset{=}{\underset{f}{-}} b)$, since this would amount to a statement like $(a \overset{=}{\underset{F}{-}} b) \equiv (\exists f)\ (a \overset{=}{\underset{f}{-}} b)$ and implication only goes from left to right here. What relativism must settle for, then, is the vacuous $(\exists f)\ (a \overset{=}{\underset{f}{-}} b) \equiv (\exists f)\ (a \overset{=}{\underset{f}{-}} b)$, and the fact that it cannot offer anything more interesting than this makes it suspect. A similar result occurs when it tries to provide an analogue of $(\forall x)\ (\forall y)\ [(x = y) \equiv (\exists f)\ (x \overset{=}{\underset{f}{-}} y)]$.

(ii) A relativist cannot do justice to "$a = b$ because $a \overset{=}{\underset{F}{-}} b$." For him such a statement must amount either to the incoherent "$a \overset{=}{\underset{F}{-}} b$ because $a \overset{=}{\underset{F}{-}} b$" or to the false "$a \overset{=}{\underset{F}{-}} b$ because $a \overset{=}{\underset{G}{-}} b$" except in a case where the meaning of '*F*' and '*G*' dictates that if two things are the same *G*'s they must be the same *F*'s—e.g., as in "$a \overset{=}{\underset{\text{cow}}{-}} b$" and "$a \overset{=}{\underset{\text{animal}}{-}} b$." But intuitively it seems that "$a = b$ because $a \overset{=}{\underset{F}{-}} b$" can be free from incoherence and falsity in a broader range of cases than this.

(iii) Relativism gives the reflexivity of identity a rather curious shape. A relativist might regard $(\forall x)\ (x = x)$ as semantically equivalent to $(\forall x)\ (\exists f)\ (x \overset{=}{\underset{f}{-}} x)$. There is nothing intrinsically counterintuitive about the latter principle, but it does seem to be derived from the conjunction of $(\forall x)\ (x = x)$ and $(\forall x)\ (\forall y)\ [(x = y) \equiv (\exists f)\ (x \overset{=}{\underset{f}{-}} y)]$ and therefore the proposed equation is curious. Alternatively, a relativist might equate $(\forall x)\ (x = x)$ with $(\forall x)\ (\forall f)\ [f x \supset (x \overset{=}{\underset{f}{-}} x)]$. But in this case the proposed equivalent is suspect because

$(\forall x)\ (\forall y)\ \{(x = y) \equiv (\forall f)\ [(fx\ \&\ fy) \supset (x \overset{=}{\underset{f}{-}} y)]\}$

is unacceptable. We must replace the latter principle with

$(\forall x)\ (\forall y)\ \{(x = y) \equiv (\forall f)\ [(fx\ \&\ fy) \supset (\exists w)\ (\exists z)\ [(x \overset{=}{\underset{f}{-}} w)\ \&\ (y \overset{=}{\underset{f}{-}} z)\ \&\ (w = z)]]\}$

We can then combine this with $(\forall x)\ (x = x)$ to get

$(\forall x)\ (\forall f)\ (fx \supset (\exists w)\ (\exists z)\ [(x \overset{=}{\underset{f}{-}} w)\ \&\ (x \overset{=}{\underset{f}{-}} z)\ \&\ (w = z)])$.

Replacing $(w = z)$ with $(\exists g)\ (w \overset{=}{\underset{g}{-}} z)$, we then get

$(\forall x)\ (\forall f)\ (fx \supset (\exists w)\ (\exists z)\ (\exists g)\ [(x \overset{=}{\underset{f}{-}} w)\ \&\ (x \overset{=}{\underset{f}{-}} z)\ \&\ (w \overset{=}{\underset{g}{-}} z)])$

as an intrinsically respectable replacement for $(\forall x)\ (x = x)$. Once again, however, the putative semantic equivalent of $(\forall x)\ (x = x)$ is a principle which has been derived from a conjunction, one conjunct of which is $(\forall x)\ (x = x)$.

Geach's remarks suggest two further arguments in favor of relativism and he outlines a third one while discussing Leibniz's Law. The two suggested arguments are:

(1) If we do not adopt relativism, and if we are faced with a situation in which (e.g.) *a* and *b* are the same body but not the same corpora, and if we are asked to perform a counting operation through time which encompasses *a* and *b*, then we shall be forced into the uncomfortable position of having to say that *a* and *b* yield "two," since they are two different corpora, but also yield only "one," since they are one and the same body. The only acceptable solution is to relativize the counting procedure.

(2) In calling a baby and an adult by the same name, we imply that they are the same human being. We cannot without qualification say they are the same individual and thereby justify calling them by the same name:

> if an individual is presented to me by a proper name, I cannot learn the use of the proper name without being able to apply some criterion of identity; and since the identity of a thing always consists in its being the same *X*, e.g., the same *man*, and there is no such thing as being just "the same," my application of the proper name is justified only if (e.g.) its meaning includes its being applicable to a *man*.[6]

---

[6] P. T. Geach, *Mental Acts*, *op. cit.*, p. 69.

Reply to (1): We should not overlook the fact that the terms of reference of given counting operations are variable. For example, even within an identity statement we can say "Those two (three, four . . .) things are one and the same" and one kind of counting operation might be to determine how many things an identity claim seeks to identify, even though, if the claim is true, it in a sense picks out only *one* thing. Alternatively, the principle governing a counting operation might be such that two things would count as one so long as there was *some* value of *f* for which we could say that they are the same *f*; otherwise they would count as more than one. In that case if *a* and *b* are the same body but not the same corpus, they would count as one. On the other hand, the operation might be such that for certain purposes *a* and *b* would count as two, because they are different corpora, whereas for other purposes they would count as one, because they are the same body. The results of a counting operation with respect to *a* and *b* will therefore vary as the operation's terms of reference vary. But there is nothing in this fact to lend support to relativism. All that follows is that not every counting operation through time will yield a correct count of different individuals and there is nothing strange about that. Nor does this imply that in counting different individuals through time, one may simply count individuals and not specific individuals, such as men, cats, chairs, etc. One cannot avoid counting specific individuals. But what one can do is to say, e.g., "Here are two corpora which are not the same under any other sortal concept, and therefore here are two individuals." This satisfies the demand that counting must be done under sortal concepts, and yet preserves a role for a non-relative diversity statement like "Here are two individuals."

Reply to (2): It is true that the use of the same name to pick out an adult and a baby through time cannot be ultimately *justified* simply by saying "They are the same individual." It must be supported by something like "They are the same human being," something which can be tested by reference to appropriate criteria of identity. Moreover, in understanding the use of the proper name, we shall have to understand how to justify that use. But this does not mean that in using the proper name we must always imply that the given individuals are both human beings (or some other specified so and so's). For example, someone might tell us that a given baby, born in 1935, and a given adult, living in 1970, are the same individual and therefore bear the same name, "John," without telling us why such a claim is true. If we believe our informant because he is generally reliable and use "John" to refer to the baby and the man, we shall not be using "John" in a way which implies that the baby and the man are the same human being, precisely because we do not know whether their being the same human being provides the justification for calling them both "John." For all we know, it might be because they are the same organic structure, or the same collection of molecules, and so on. Furthermore, even in a situation in which the use of "John" does knowingly rest upon, and hence imply, their being the same human being, there is nothing wrong with suggesting that the use of the proper name *also* implies their being the same individual. The truths in (2) can be preserved, then, without going on to embrace relativism.

Geach offers a further argument for relativism by trying to undermine Leibniz' Law (See "Identity"). Strictly speaking, in so doing he is trying to establish more than just relativism as it is presented in this paper, since his doctrine is intended to cover the "contemporaneous" identity statements as well as statements concerning identity through time. I shall first sketch his argument as it occurs in its more comprehensive form and then indicate how it might be revised to give it a suitably limited scope.

The more comprehensive argument is: If identity is non-relative, then there must be a criterion of identity which may be expressed as "Whatever is true of something identical with an object *a* is true of *a*, and conversely" or as "*x* is identical with *y* only if whatever is true of *x* is true of *y*, and conversely." Call it the "Indiscernibility of Identicals" and label it 'II.' II is unacceptable as it stands because it uses "is true of" in an unrestricted way. That is, it doesn't restrict "is true of" to predicables of some definite language *L* and is therefore vulnerable to the semantic paradoxes. But if we restrict "is true of" and attach the rider "in language *L*" to II, then one of two equally unpalatable results occurs. Either *a* and *b* may be identical, since indiscernible, in one language (*L*1) but not identical, since discernible, in another language (*L*2), even though *L*2 contains *L*1 in the sense that *L*2 is simply an enriched extension of *L*1; or, if the variables in *L*1 and *L*2 are not regarded as ranging over the same values, because identity conditions in the two languages differ, then a proliferation of entities results which invites the use of

Occam's razor. To take an example used by Geach, say $L1$ contains a concept "surman" which behaves such that if two men have the same surname, then they are the same surman. And say that $L1$ contains no concept which could be used to distinguish given surmen. Also, say that $L2$ contains concepts which can be used to distinguish given surmen. In that case, according to the modified version of II, any two surmen may, relative to $L1$, be the same individual, since they will be indiscernible in $L1$, whereas relative to $L2$ not every two surmen can be the same individual, which is curious. Alternatively, if, because of the difference in identity conditions, the constants in $L1$ cannot be regarded as picking out the same individuals as those in $L2$, then the mere introduction of an additional set of concepts will be regarded as introducing an entirely new set of entities, which offends Occam's spirit.

The limited argument is: Say $L1$ contains the concept "corpus" but not the concept "body," as these concepts have been explained above. And say there is a situation in which, through time, a and b are the same body but different corpora because not all their parts are the same. A user of $L1$ is unable to say that a and b are the same body and suppose that $L1$ contains no other concepts under which a and b might be judged identical—i.e., in $L1$ there is no concept 'f' such that $(\exists f)$ $(a \underset{f}{=} b)$. In that case, by the modified II, a and b are different in $L1$ because there is no sortal concept under which they can be identified, and yet they are the same in any language enriched by the mere addition of "body." Alternatively, if 'a' and 'b' must pick out different individuals in $L1$ from those picked out in the richer language, then an intolerable entity proliferation results.

Geach's argument has been criticized by Fred Feldman and John Perry.[7] Among other things, Feldman claims that there is nothing particularly unattractive about making identity language-relative, but that even if one avoids this by making constants and variables in different languages range over different sets of individuals, proliferation need not result because the richer language is not tied to the existential commitments of the poorer one. Perry argues that either Geach's purported identity statements are not really identity statements—e.g., "a and b are the same word-types" amounts to "a and b are the same type of word"—or the constants they use do fail to introduce the same particulars in different languages, but that in the latter case there is no undesirable entity proliferation. I think that Feldman and Perry are right on the proliferation question to the extent that if the constants in different languages do introduce a different set of entities, then, even if two languages are related as richer to poorer, the existential commitments of the two languages do differ, simply because the individuals they introduce differ. It is plausible to think otherwise only if one thinks that in using the richer language, someone is also thereby using the poorer language which it is said to "contain." But this is a mistake. Someone cannot be using both languages at once. Consequently, contrary to what Geach suggests, the thesis that the constants in different languages pick out different individuals does not imply that if statements in the richer language are true, then statements in the poorer "contained" language must also be true, since in this context the question of truth cannot be separated from the question of existential commitment. For example, if A uses $L1$ and B uses $L2$, and A and B both say "a and b are different corpora," then, according to the given hypothesis, if what B says is true, what A says is not true, because the existential implications of what A says are false. At the same time, although the charge of entity proliferation will not hold up, the view which Geach is attacking *is* subject to the charge that it makes the set of individuals over which the variables and constants of a language range a function of the sortal concepts contained in that language and rules out the possibility of successful communication between someone who uses a poorer language and someone who uses a richer language. Thus, someone who uses $L1$ cannot understand someone who uses $L2$ when ostensibly each is talking about the same individual corpus—e.g., when the first person says "This corpus is brown" and the second says "Yes, it is brown." This is intolerably restrictive. Their ability to communicate at this level does not seem to depend upon their both being able to say of a given corpus that it is the same individual as such and such a corpus.

Feldman's suggestion that there is nothing unattractive about making identity claims relative to a language is itself not very attractive and Geach is right not to be convinced by it. Feldman suggests a more promising line of argument when he suggests that (1) even if II must be abandoned as a mark of absolute identity, this does not require abandoning absolute identity, and (2) although a version of II which is not relativized to a language may *seem* to create semantic paradoxes, appearances may be deceiving and the paradoxes might be resolvable without relativizing "is true of." An even more convincing objection, however, would be to point out that if an unrestricted II does generate semantic paradoxes, the restriction on II which is needed to avoid this does not have to be as severe as Geach thinks. We could say, e.g., that *a* and *b* are identical only if they are indiscernible in all languages of the order *ab*, where a language is of order *ab* if it ranges over just those items of the same order as *a* and *b*. Thus, in the "surman" example, although two surmen may be indiscernible in *L*1, they are discernible in *L*2 and, since *L*2 is of the appropriate order, they are therefore not absolutely identical. In the "corpus" example, although in *L*1 there is no concept under which *a* and *b* can be judged identical, there is such a concept in *L*2, and therefore they are absolutely identical. *All* things considered, then, Geach fails in his attempt to undermine non-relative identity by undermining Leibniz' Law.

---

[7] See Fred Feldman, "Geach and Relative Identity," *Review of Metaphysics*, vol. 22 (1968–69), pp. 547–555; and John Perry, "The Same F," *Philosophical Review*, vol. 79 (1970), pp. 181–200.

David Wiggins sets himself against both Geach and a Lockean by rejecting thesis (b).[8] Since he labels the thesis '*R*', I shall use his label in the remainder of this paper. His argument against *R* has three parts:

(i) *R* is incompatible with Leibniz' Law—in symbols, $(a = b) \supset (\forall \phi)(\phi a \equiv \phi b)$. To conform to Wiggins' phraseology, label the law 'LL' in this context.

(ii) LL cannot reasonably be abandoned.

(iii) Specific cases, real or imaginary, which purportedly support *R* can be redescribed in such a way that they no longer offer such support.

Wiggins thinks the most direct way of defending (i) is to take "$a \frac{=}{G}$" as a value of ϕ in LL. If we substitute in a straightforward way, then we get

$$(a = b) \supset [(a \tfrac{=}{G} a) \equiv (a \tfrac{=}{G} b)] \qquad (1)$$

R insists on the possible truth of the conjunction

$$(a \tfrac{=}{F} b) \;\&\; (a \tfrac{\neq}{G} b) \;\&\; (Ga \;\&\; Gb) \qquad (2)$$

The first conjunct of (2) yields "$a = b$" which with (1) yields

$$(a \tfrac{=}{G} a) \equiv (a \tfrac{=}{G} b) \qquad (3)$$

Since "$Ga \supset (a \frac{=}{G} a)$" is an axiom, therefore (3) combined with (2)'s third conjunct yields

$$a \tfrac{=}{G} b \qquad (4)$$

And (4) clashes with the second conjunct in (2). Consequently, the conjunction of LL and *R* yields a contradiction.

He defends (ii) by drawing on three premisses:

(1) LL marks off what is peculiar to identity. Features like transitivity, symmetry, and reflexivity are unable to do this, either individually or jointly.

(2) There is no independent case against LL. Appeals to non-extensional contexts are ineffective for want of a genuine reference to the individual in question.

(3) If LL is abandoned, then no weaker replacement is available to underwrite legitimate substitutions of identicals.

Consider Wiggins' defense of (i). If "$a \frac{=}{G}$" counts as a value of ϕ in LL, then the question arises of whether LL must yield (1), or whether it should yield something like

$$(a = b) \supset \{(a \tfrac{=}{G} a) \equiv [(a \tfrac{=}{G} b) \vee (\exists x)((a \tfrac{=}{G} x) \;\&\; (x = b))]\} \qquad (5)$$

In the vernacular (5) says that if $a = b$, then *a* is the same *G* as *a* if and only if either *a* is the same *G* as *b* or there is an individual *x* such that *a* is the same *G* as *x* and $x = b$. The first conjunct of (2) combined with (1) yields

$$(a \tfrac{=}{G} a) \equiv [(a \tfrac{=}{G} b) \vee (\exists x)((a \tfrac{=}{G} x) \;\&\; (x = b))] \qquad (6)$$

and (6) combined with (2)'s third conjunct and

"$Ga \supset (a \frac{=}{G} a)$"

yields

$$a \tfrac{=}{G} b \vee (\exists x)((a \tfrac{=}{G} x) \;\&\; (x = b)) \qquad (7)$$

And unlike (4), (7) does not clash with anything in (2). Hence, if (5) is what LL yields and not (1), then there is no contradiction in the conjunction of LL and R.

---

[8] See David Wiggins, *Identity and Spatio-Temporal Continuity* (Oxford, 1967), Parts 1–3.

Unless we adopt a tenseless interpretation of the notation being used here, LL should be regarded as yielding (5) and not (1). Otherwise, LL would be at a loss to account for identity through time in situations where the persisting individual undergoes a change. For example, say a and b are the same cup at consecutive times such that the cup is orange ("a is orange") and was non-orange ("b was non-orange"). If we take "is orange" (O) as a value of ϕ in LL and perform a straightforward substitution of the kind Wiggins does with "$a \underset{G}{=}$," then we shall get

$$(a = b) \supset (Oa \equiv Ob) \tag{8}$$

Now, either "Ob" is a tensed statement or it is not. If it is tensed, then either it is incoherent because 'b' picks out the cup as it existed initially or "Ob" really means "b was orange," in which case the condition that b was non-orange cannot be satisfied. In other words, substituting in LL in this manner dictates either that the conjunction "$(a = b)$ & (Oa)" generates something incoherent or that the conjunction guarantees that the given individual was initially orange. Both alternatives are unacceptable.

To avoid this dilemma when dealing with identity through time, one must substitute in LL as follows:

$$(a = b) \supset \{Oa \equiv [Ob \vee (\exists x)\ (Ox \ \&\ (x = b))]\} \tag{9}$$

This in effect is to make clear that if $a = b$ and Oa, then the indiscernibility condition is satisfied either if Ob or if b has now become an individual which is orange. And this is exactly what was done above in the case of "$a \underset{G}{=}$" when LL was made to yield (5) instead of (1).

Now, adopting this type of substitution procedure *is* curious to the extent that it involves interpreting "$\equiv \phi b$" in LL in such a way that it is a disguised disjunction. There are two ways of avoiding the need to do this. One way is to insist that the formulae here are to be treated tenselessly. In (8), then, 'Ob' is to be understood such that it is true if either "b is orange" or "b was orange" or "b will be orange" is true. In that case, there is nothing incoherent about "Ob" and its truth does not entail that b was orange. If we abide by this interpretation, however, then the friction between LL and R disappears in a different way. For, "$(a \underset{G}{=} b)$ & $(a \underset{G}{\neq} b)$" is no longer inconsistent. The first conjunct can be true as a result of the truth of "a has been the same G as b" and the second one can be true as a result of the truth of "a is not the same G as b," and there is nothing to prevent both statements being true (cf. "Ob" being true because "b will be orange" is true and "$\sim Ob$" being true because "b was not orange" is true). The only way that friction can develop between "$a \underset{G}{=} b$" and "$a \underset{G}{\neq} b$" is by reading the second statement as "a neither has been, nor is, nor will be, the same G as b," and then it is not equivalent to "$a \underset{G}{\neq} b$" *as the latter occurs in R.*

Before considering the second way of removing the need for performing curious substitutions in LL when dealing with identity through time let me deal very briefly with one possible objection. Someone might argue that in (8) "Ob" can be tensed and read as the coherent "b is *now* orange." Hence the putative dilemma is resolved without such an intricate substitution procedure and without resorting to a tenseless interpretation. Two things are wrong with this objection: First, if "Ob" is read in this manner, then the same reading should apply to "Oa" and again incoherence would result. Second, even if the need for an intricate substitution procedure is thereby eliminated in the case of "O," it would not help the contention that LL and R are inconsistent. For, applying LL to get (1) and combining (1) with (2) then merely yields the conclusion that b is *now* the same G as a, which hardly contradicts "a is not the same G as b" as it occurs in (2).

The second way of eliminating the need for curious substitutions in LL when dealing with identity through time is, not too surprisingly, to modify LL to make it look like

$$(a = b) \supset (\forall \phi)\ \{[(\phi a \supset (\exists x)\ (\phi x \ \&\ (x = b))] \ \&\ [\phi b \supset (\exists x)\ (\phi x \ \&\ (x = a))]\}$$

In the vernacular it would read "If a is identical with b, then for any ϕ if a is ϕ, then there is an individual which is ϕ and which is identical with b, and if b is ϕ then there is an individual which is ϕ and which is identical with a." Thus, if "$(a = b)$ & $(a \underset{G}{\neq} b)$" is true, although "$\underset{G}{\neq} b$" is true of a and not of b, in being true of a it is true of an individual with which b is identical. Granted, modifying LL in this way completely trivializes it, but perhaps this is what one must do if LL is to accommodate identity through time and lend itself to straightforward substitutions.

The modified, trivial version of LL will satisfy two of the conditions which Wiggins thinks must be satisfied either by LL or by some replacement. It provides a feature of identity not shared by other transitive, symmetrical, and reflexive relations and it does not come to grief over non-extensional contexts. It will not satisfy Wiggins' third condition, however, since it does not underwrite a straightforward substitution of identicals even in cases where such a substitution might appear legitimate. For example, it will not warrant moving from "My infant body and current body are identical and my current body is now missing a gall bladder" to "My infant body is now missing a gall bladder." All it will underwrite is the move to "My infant body is identical with something which is now missing a gall bladder." Yet at least it does underwrite this much, and it is arguable whether LL should underwrite anything more ambitious than this if it is intended to be a completely unrestricted principle. For, although LL can accommodate intentional contexts, an unrestricted substitutivity principle cannot. For example, "$(a = b)$ & (S thinks that a is F)" does not entail "S thinks that b is F," whereas "$(a = b)$ & (S thinks that a is F)" does entail "There is an individual such that S thinks that that individual is F and that individual $= b$." Consequently, as far as substitutivity is concerned, we must impose certain restrictions in view of intentional contexts, and once we do this, then the door is open for adding a restriction regarding identity contexts like "$a \frac{=}{G}$." Moreover —although I think this point is really unnecessary —it is not clear why there *must* be a general principle with a shape like LL's to underwrite the legitimacy of those inferences involving the substitution of identicals which are legitimate.

Wiggins' attempt to use LL against R is therefore unsuccessful. There is no genuine clash between LL and R unless LL is construed as an unlimited substitutivity principle—and then it is no longer plausible. I shall not bother to supplement this general discussion with an attempt to show that Wiggins' interpretations of specific cases are mistaken. I shall leave this for the reader to decide, since, as Wiggins himself suggests, the real battle must be fought at the level of principles.

John Perry, in criticizing Geach on Frege's behalf, attacks R. Perry tries to show that "$(a = b)$ & $(Fa$ & $Fb)$" entails "$a \frac{=}{F} b$" and that "$(a \frac{=}{F} b)$ & $(Ga$ & $Gb)$ & $(a \frac{\neq}{G} b)$" is inconsistent. One of his arguments against R is as follows:

> If "$(a \frac{=}{F} b)$ & $(Ga$ & $Gb)$ & $(a \frac{\neq}{G} b)$" is consistent, and if we make substitutions in the third conjunct which are licensed by the first conjunct, then we shall derive the absurd "$a \frac{=}{G} a$". Alternatively, if we refuse to make the substitution, we deny the substitutivity of identicals, and this is equally absurd.[9]

Geach can dismiss this argument by insisting that the substitution of identicals is permissible only if the value of f in "$a \frac{=}{f} b$" remains constant. But an advocate of R who rejects thesis (a) does not have this reply open to him. He does have a different reply available, however. He can say that a straightforward substitution is not allowed here precisely because the context concerns identity through time. The most he will allow is either that "$a \frac{\neq}{G} a$" might be read as the non-absurd "a is not the same G as it once was," or that we are simply deducing "a is not the same G as an individual with which a is identical." If, as Perry suggests, this involves abandoning a "traditional and rather plausible" logical doctrine, then the doctrine's plausibility is deceptive and the tradition is in need of a change.

Like Wiggins, Perry also attacks R by considering a case which ostensibly supports the consistency of "$(a = b)$ & $(Fa$ & $Fb)$ & $(a \frac{\neq}{F} b)$" and offering a redescription which makes things look otherwise. For instance, say a statue of a man is molded from a piece of clay and later the same piece of clay is molded into a statue of a horse. Call the two statues $S1$ and $S2$. An advocate of R will say "$S2$ is the same piece of clay as $S1$ but not the same statue." According to Perry, on the other hand, either $S1$ and the piece of clay are not identical or "$S1$ and $S2$ are the same piece of clay" is not really an identity statement but has the force of "*S*1 and *S*2 are made of the same piece of clay." Now, I'm not too concerned about refuting Perry's analysis of this particular example, since if the example fails to serve *R*'s ends, then more effective ones can be found. What I am primarily interested in are the reasons Perry offers on behalf of his analysis, since a critic of *R* must win the battle at the level of principles if he hopes to convert anyone at the level of specific cases. The merit of Perry's position turns on whether he supports his first disjunct effectively, viz., the claim "*S*1 and the piece of clay are not identical." He admits that the claim may look paradoxical to some but argues that appearances are deceiving because "there are things true of the one not true of the other (for example, the piece of clay was bought in Egypt in 1956, but not the statue) and the piece of clay may remain with us long after the statue is destroyed."[10]

---

[9] John Perry, *op. cit.*, p. 186.

Perry's argument draws on LL in a way which makes the law insensitive to the fact that we can say things like "The statue was bought in Egypt in 1956, although it wasn't a statue then; it was just a piece of clay" and "The statue may remain with us long after it ceases to be a statue, as a piece of clay." And even if this objection is unacceptable, Perry must still come to grips with the version of LL according to which the piece of clay and the statue are identical in so far as the statue is identical with an individual bought in Egypt in 1956 and an individual which may remain with us long after the statue is destroyed as such. Like Wiggins, then, Perry fails to show that cases described in accordance with *R* are incompatible with LL, in any sense in which the latter law is tenable. Also like Wiggins, he is right in claiming that *R* rules out the unlimited substitution of identicals, but this should be regarded as a merit rather than a demerit.

Geach's theory is correct in so far as it contains *R* (thesis (b)) but mistaken in so far as it contains thesis (a). Wiggins' theory is correct in rejecting Geach's theory, but incorrect in rejecting *R*. The Lockean theory is correct in rejecting thesis (a) and in advocating *R*. All three theories are correct in accepting thesis (c).

University of Guelph

Received October 27, 1970

Appendix

John Locke undeniably accepts *R*, though there are two reasons for thinking that he would reject thesis (a).

(i) He introduces the idea of identity as the idea we frame when we compare the "very being" of a thing by comparing the thing as it exists at one time with itself as it exists at another time, but there is no suggestion that, in so doing, we must be thereby thinking of the given thing as being the same so and so as itself at another time.

(ii) Locke claims that "one thing cannot have two beginnings of existence, nor two things one beginning" (*Essay*, II xxvii 1), a claim which can be symbolized as

$$(\forall x)\ (\forall y)\ \{(x = y) \equiv (\exists w)\ (\exists z)\ [(x = w)\ \&\ (tpw)\ \&\ (y = z)\ \&\ (tpz)\ \&\ (Szw)]\}$$

where '*tp*' stands for "begins at time *t* and place *p*" and '*S*' stands for "is of the same sort as." For him it is a claim which we can consider before considering

$$(\forall x)\ (\forall y)\ [(x = y)\ \equiv (\exists f)\ (x \overset{=}{\underset{f}{-}} y)]$$

Consequently, it is gratuitous to think that for him the first claim has a structure like

$$(\forall x)\ (\forall y)\ (\exists f)\ \{(x \overset{=}{\underset{f}{-}} y) \equiv (\exists w)\ (\exists z)\ [(x \overset{=}{\underset{f}{-}} w)\ \&\ (tpw)\ \&\ (y \overset{=}{\underset{f}{-}} z)\ \&\ (tpz)\ \&\ (Szw)]\}.$$

---

[10] John Perry, *op. cit.*, pp. 198–199.

American Philosophical Quarterly
Volume 9, Number 1, January 1972

IV. SEMANTICS FOR LOGICAL AND NOMIC MODALITIES

NICHOLAS LA PARA

I

THAT law statements are in some way different from accidental generalizations is not controversial, but the explanation and analysis of that difference is a matter of dispute. Some, including the author, believe that law statements are logically different from accidental generalizations, and that modalities distinct from the logical modalities are employed in law statements. Many more philosophers think there is no logical difference between law statements and accidental generalizations. These modern Humeans think that laws differ from other generalizations only in that we *opt* to treat laws in a way systematically different from the way we treat the other generalizations.

Either group, it seems, should be interested in a semantical theory—and hence, a logical system—of nomic modalities. The first group, insofar as the system can be applied to analyze the concepts grouping around laws—conditionals, dispositions, hypothetical reasoning, etc.—will regard it as a vindication of their view, especially as the system illuminates the difficult concepts of possible individuals, possible alternative states of this world, and possible alternative worlds to this one. The Humeans, since they admit that the treatment accorded to "lawful" generalizations is systematic, may regard the system of nomic modalities as a rather oblique way, not to be taken literally, of seeing into that systematicity. But even so regarded, such a system would be valuable.

The full range of applications cannot be presented here, for a good deal of discussion is required to see what features the applications must preserve if they are to be adequate analyses of such resistant concepts as, for example, subjunctive conditionals. Some indications of applications will be given, however.

II

A successful system of putative nomic modalities must meet certain requirements. First, and most basic, it must be shown that nomic modalities are distinct from the logical ones. The present system therefore contains both, so that the interrelationship of the modalities can be clearly seen.

Since laws obviously contain quantifiers, we must also have a semantics for a combined system of quantifiers and modalities. In this context we have to face the issue of the "referential opacity" of logical and nomic modalities, which might call into question the coherence of such a combined system. Since the issue of "referential opacity" is generated by arguments centering around identity and the use of definite descriptions, the system is made rich enough to contain these. A complete resolution of the problems of quantifying into and substituting into modal contexts is provided. We then also have the basis of a theory of definite descriptions, one that does not analyze them away but takes them as linguistic units in their own right. The logical differences between names and definite descriptions are consequently clarified.

There are (in an ordinary and harmless sense of "are") other possible worlds, such as the worlds of Frankenstein, Superman, or Pegasus. The laws of those worlds are very different from ours. Our laws do not apply to the individuals just mentioned, whose very concept involves the kind of possible world containing them. But the laws of our world do apply to more than the actually existing individuals. They apply to merely possible ones as well—Hamlet, the secretary in my lap, the event of impressing a voltage across this wire rather than, as now, dissolving it in acid. Another way of acknowledging this is to point out that alternative possible states of our world (the world embodying the true laws) can contain individuals that don't exist in this, the real state. Our system must therefore permit reference to and quantification over these individuals.

Sharp metaphysical problems face the concepts of possible individuals and possible worlds. One is the nature of the reality of those entities. Another is the criterion of identity for individuals from state to state, and world to world. If alternative states and worlds are to provide the basis for discourse about "might-have-beens" and "could-be's," then some individuals in those other states and worlds must be identical, and some non-identical, with the actual individuals of the real state. Furthermore, a language permitting reference to possible individuals creates the semantical problem of providing grounds for the truth-values of statements about those individuals.

The metaphysical problems are resolved, first, by adopting a language containing non-referring names. These are taken to be the names of merely possible individuals, although they don't really refer to anything. Reference to merely possibles proceeds by use of non-referring names. Use of "merely possible individual" is a perhaps misleading meta-idiom. The second step in meeting the metaphysical difficulties is taken by constructing possible worlds, etc.,—or more precisely, entities which permit idiomatic talk in terms of possible worlds and so forth—out of sets of linguistic items. After some discussion, this is done in complete detail in Sect. VII.

This method of construction provides the groundwork for a criterion of trans-state and trans-world identity of individuals. Roughly, certain structures containing linguistic items are set up to play the role of possible worlds and possible states of a world, and in these, sentences play the role of facts, and names play the role of individuals. There is of course no problem with identifying names across these "possible worlds," and from a nominalistic point of view, this gives the identity criterion for individuals. This, by the way, indicates that if $b_1 = b_2$ holds (fails) in one state of one world, it should hold (fail) in every state of every world.[1] Independent arguments for this feature are given below, and it will be incorporated in the construction.

What of the semantical problem of providing truth-grounds for sentences about possibles—that is, sentences containing non-referring names? Following van Fraassen,[2] I allow that while there are only two truth-values, still certain sentences may lack a truth-value entirely (this is not the same as giving them a different, third truth-value, and should not be treated as such). In particular, this may happen to sentences containing non-referring names. *But it need not.* The way is open for independently specifying sufficient conditions for the truth or falsity of such sentences. Thus, the present system is constructed so that sentences having logically valid forms are true, indeed necessary, *whether or not they contain non-referring names.* We can also require that any sentence implied by a true law of nature be true; a further word on that is provided in Sect. IX. In applying the system we could also demand, for example, that sentences about fictional characters which accord with a certain standard fictional account also be true.

III

Some requirements on a system that can provide an analysis of nomic notions have been laid out, and the method of meeting those requirements sketched in a preliminary way. Certain results of general logical and semantical interest have been seen to flow as by-products of the method employed. There are others. One is that the system is constructed so that the logic generated applies to empty domains. The logic is therefore completely free of existential assumptions. Moreover, we have a semantical foundation for this very rich system of free logic, which settles questions of interpretation for empty domains.

A second result comes about through the devices used to resolve the problem of the "referential opacity" of modal contexts. The leading idea is to distinguish between modalities *de re* and *de dicto*, and to show that names and definite descriptions behave differently in these contexts. The sentential operator notion of modality (i.e., the *de dicto* sense) is taken as basic, and with the help of the λ-operator, *de re* modality can be defined—in fact *de re* modality is seen to be a special case of a more general use of modality. The λ-operator produces predicates out of open formulas; $(\lambda x, y)(Rx, y)$ is just the two-place predicate R. With this operator one can then define predicates which contain modal operators; for example

$$(\lambda x)\ (\Box Fx)$$

is the predicate "being-necessarily-F." With such predicates, sentences using modality *de re* (which is taken to be "modality in the predicate") can be formed. In this way we obtain a theory of the two types of modality. This approach, incidentally, sees it as a mistake to consider, as some have done, that modality inside a quantifier is *ipso facto* modality *de re.*[3]

---

[1] Note that I dispense with quote and quasi-quote conventions where confusion seems unlikely.

[2] Bas van Fraassen, "Singular Terms, Truth-value Gaps, and Free Logic," *The Journal of Philosophy*, vol. 63 (1966), pp. 481–495.

IV

A leading idea in understanding the logical modalities has been that necessary truth is truth in all possible worlds. Carnap has applied this idea in a nominalistic way in *Meaning and Necessity*. Can an analogous idea be mobilized to help with the explication of nomic modalities?

It will be helpful at this point, for reasons that will become apparent presently, to drop for a bit the term "world." In Carnap's construction, the role of possible world is filled by state descriptions. Though we shall deal with entities analogous in certain ways to state descriptions, let us call our entities "histories."

A state description characterizes completely, to the limits of the primitive predicates and names of the language, a logically possible state of affairs (not necessarily at one instant of time; the language might allow for mentioning times and dates). We should like our histories to do the same. In a sense then, histories describe the "basic" possible facts. This does not mean that non-modal statements formed solely in terms of primitive predicates and extensional connectives do not have modal implications. They do. And we could arrange for further such implications by putting restrictions on histories (and on later entities to be constructed) in much the same manner as Carnap uses meaning postulates to put restrictions on state descriptions. (But to simplify the exposition, I shall generally ignore this latter possibility.) So, a history is a basic level of description in a syntactic sense, and in another as well, to be noted in Sect. X. (Of course, we would expect a useful language to have as primitive predicates which are basic in a semantical and epistemological sense.)

If the histories represent alternative logically possible states of affairs, we might expect that a statement is nomically necessary if and only if it holds in all members of a subset of the set of histories, where that subset is considered to "correspond" to actuality. The histories in that subset represent the *nomically* possible alternative histories. Other subsets represent other notions about what is nomically possible. It would be suitable, then, to call each such subset a *world*, where worlds now represent logically possible determinations of what is nomically possible.

But a problem arises. We can no longer identify logical necessity with holding in every history. The reason is that undoubtedly there will be logically necessary statements in which nomic modalities appear (nomic statements). But for nomic statements we cannot give a criterion for holding *in a history*, since the same history appears in different worlds. Whether a nomic statement holds depends not just on a given history, but on the alternatives to that history; that is, on which world we are considering.

This problem arises because, in following up the obvious ideas related to nomic modality, we are led to introduce more structure into our constructions than usual. We now have a number of "levels"—statements, histories, worlds, the set of all worlds, etc. We have seen the difficulty in trying to tie logical modality to just one level.

The solution is simple. We must first recognize that unlike Carnap's state descriptions, the entities which in this construction represent alternative possibilities must show hierarchical structure. That is, when laws as well as "matters of fact" are taken into account, reality may correspond to world w with one of its histories h representing the actual matters of fact, or to world w' with one of its histories h' representing the actual matters of fact. Call such suitably structured entities "world structures." Then a statement $\Box p$, where $\Box$ represents logical necessity, holds in a world structure W just if p holds in every world structure. And a sentence Np, where N represents a concept of nomic necessity more general than that mentioned previously (since p need not be a sentence whose holding relates only to histories), holds in a world structure W just if p holds in every world structure containing the same world as W.

Now, how will histories differ from Carnap's state-descriptions? Like Carnap, a history will contain one, but not the other, of each pair of basic sentences. Also, in the interest of a simpler definition of holding, we will include the appropriate truth-functional compounds of the sentences in the history, i.e., the compounds validly implied by the sentences in the history. Moreover, since identity is one of the primitive predicates, we must place suitable restrictions on the contents of the history (see clause (5) of the definition of a history as given below).

But we have not yet taken account of the fact that our language permits non-referring names.

---

[3] See, for example, A. N. Prior, *Formal Logic*, 2nd ed. (Oxford, 1962), p. 209; and G. E. Hughes and M. J. Creswell, *An Introduction to Modal Logic* (London, 1968), p. 183.

This fact means that we must acknowledge that alternative states of affairs can differ not only in which individuals have which properties, but in which individuals exist. Such existence facts must be represented in our histories. To this end, we introduce into the language "basic existence statements," which fall into pairs such as "$(\exists x)$ $(x=b)$," "$\sim(\exists x)$ $(x=b)$," where b is a name. The first of these is intended to say that b exists, and the other to deny it. These statements are taken to be as fundamental as atomic statements; each history contains one but not the other of each pair of basic existence statements.

Since the rules for holding for statements of the form "$(\forall x)E(x)$" are independent of those for "$(\exists x)$ $(x=b)$," the use of the existential quantifier in basic existence statements is somewhat previous. However, virtue will triumph, for "$\sim(\forall x)\sim(x=b)$" and "$(\exists x)$ $(x=b)$" are provably equivalent.

If the use of basic existence statements is to be adequate in representing the various possibilities of individual existence, we must assume that every actual individual is named in the language (we shall have other names as well). This same assumption will allow us to interpret universal quantification in terms of substitution instances.

However, Quine has noted that in certain rich languages not every entity can be named.[4] In particular, the class of all open sentences of the language which do not satisfy themselves can't without paradox be named in a language adequate to express class membership. But languages where this problem interferes with the substitution-instance interpretation of quantification must be second-order languages. If we content ourselves with a first-order language, we avoid such problems.

But even taking this first-order stance, might there not be entities more numerous than the possible names? To this I think there are two possible responses: either we staunchly assert that there are names enough, because, say, the names are indexed by the ordinals; or in the spirit of Carnap, we can simply accept that the quantifiers are relativized to the language, that '$(\forall x)$' is not to be understood as "for any actual x," but rather "for any actual x acknowledged by name in the language"; and similarly for the other quantifiers.

In any case, in establishing the semantically valid statements, no account is taken of exactly how many names are in the language, except that it is assumed there are at least countably many.

Whether results similar to those achieved here can be obtained without the problems of substitutional quantification, and without postulating a "real" realm of merely possible entities, remains to be seen.

V

Before we construct the entities discussed we must note a certain assumption and its implications. We shall assume that the languages dealt with have a specific and constant interpretation of predicates and a specific and constant interpretation of names. This means at least that the predicates have specified extensions in the real world, and each name has either a specified referent, or none at all. But what does specificity mean for the histories we shall construct; in particular, what does it indicate concerning the inclusion in histories of identity statements of the form $b_1=b_2$, where b_1,b_2 are names?

It has already been noted that the nominalistic interpretation of possible individual requires that every history contain the same basic identity statements; for every b_1,b_2, either every history contains $b_1=b_2$, or every history contains $\sim(b_1=b_2)$. But is this reasonable otherwise?

It is not only reasonable, I think, but required by the specificity of the language, in particular by the constancy of the interpretation of names. For suppose that one history contains $b_1=b_2$, and another contains $\sim(b_1=b_2)$. Then the system would allow that while b_1 and b_2 are identical under one state of affairs, it is possible that they not be, for under another state of affairs they are not identical. This, however, represents a real shift in naming. For, consider two names, "Cicero," "Tully," with the same actual referent, and ask yourself, "Suppose Cicero hadn't been Tully?" What can you suppose but that, contrary to fact, the names had been made to stand for different individuals? Or, take two names with different actual referents, and ask, "Suppose Cicero had been Mark Antony?" One *can* interpret this as a request to imagine Cicero placed in one or more of Mark Antony's situations. But that effort at interpretation only points up the impossibility of doing *literally* what is requested, an impossibility imposed by the fact that the names have different referents. Now take two names with the same fictitious referent, say, "Alexey" and "Alyosha" in *The Brothers Karamazov*. To suppose that Alexey was not

---

[4] W. V. Quine, *Philosophy of Logic* (Englewood Cliffs, N. J., 1970), pp. 53, 92–93.

Alyosha, one must imagine the story changed so that these names are made to name different characters. Finally, try supposing that Alexey had been Dmitri, and again you must change the story by changing the name conventions in it, or be faced with the same absurdity as for "Cicero" and "Mark Antony."

It should be noted that in those cases where we seem able and are ready to make suppositions of the form described *without* assuming changes of name conventions, we are really taking a "name" to be a short way of specifying a description. Thus, "Suppose Clark Kent were not Superman," might be an invitation to imagine that someone else can leap tall buildings at a single bound, etc. This simply means that names according to grammar may not be names in function.

All the changes envisioned above are changes of convention which breach the required specificity of names by making their interpretation inconstant. Therefore it will be required, as mentioned, that every history contain the same basic identity statements. The basic identity statements to be included are chosen prior to constructing histories, for that choice is part of identifying and specifying the language.

The result of this is that $b_1 = b_2$ if true (false) will be necessarily so. This should cause no alarm. It does not mean in any *vicious* sense, as Harman has suggested,[5] that b_1 and b_2 *necessarily* name what they name. It simply means that b_1 and b_2 name what they name as part of the ground rules of the language; within those rules they must name that, and no other. But one need not play with those rules—use that language, that is.

Let us turn now to the constructive task.

VI

We must now specify the formulas of our language. First, we specify the following metavariables:

$b, b_1, \ldots$ for names
$x, x_1, \ldots$ for individual variables
$u, u_1, \ldots$ for names, definite descriptions, and variables
$F, F_1, \ldots$ for predicates, primitive and otherwise
$E, E_1, \ldots$ for formulas
$E(x, x_1, \ldots)$,
$E(b, b_1, \ldots)$ for formulas containing at least the items indicated by the variables in the parentheses.

As basic vocabulary we assume an infinite number of predicates of any number of places; the primitive predicate '$=$'; names for every existing thing (see Sect. IV), and other names besides; the connectives $\sim$, &, v, $\supset$; the quantifiers $(\forall x)$, $[\forall x]$, $(\exists x)$;[6] the operators $(\lambda x_1, \ldots, x_n)$, (ιx), $\Box$, $\mathcal{N}$; and suitable parenthesis punctuation.

Predicates, formulas, and definite descriptions are then defined by the following recursion:

1. If F is a primitive predicate, then F is a predicate.
2. $(\iota x)\, E(x)$ is a definite description.
3. If F is an n-place predicate, then $Fu_1 \ldots u_n$ is a formula; $u_1 = u_2$ is a formula.
4. $\sim E$ is a formula.
5. E_1 & E_2 is a formula.
6. $E_1 \vee E_2$ is a formula.
7. $E_1 \supset E_2$ is a formula.
8. $\mathcal{N}E$ is a formula.
9. $\Box E$ is a formula.
10. $(\forall x)E$ is a formula.
11. $[\forall x]E$ is a formula.
12. $(\exists x)\,(x = b)$ is a formula.
13. If there are no free variables in $E(x_1, \ldots, x_n)$ other than x_i, $1 \leqslant i \leqslant n$, then $(\lambda x_1, \ldots, x_n)E(x_1, \ldots, x_n)$ is an n-place predicate.
14. The above is a complete specification of predicate, formula, and definite description.

VII

It was suggested earlier that a history will be a set of sentences. First we choose the basic identity statements that will be in all histories; call the set of them the ***basic identity set***. Then a history contains

1. The sentences of the basic identity set
2. Exactly one atomic sentence from each basic pair
3. Exactly one basic existence statement from each basic pair of those

---

[5] Gilbert H. Harman, "A Nonessential Property," *The Journal of Philosophy*, vol. 67 (1970), pp. 184–185.

[6] The system contains two universal quantifiers, $(\forall x)$ ranging over "actuals," and $[\forall x]$ ranging over "actuals" and "possibles"; $(\forall x)E(x)$ will be true exactly when for every name b *of an actual thing*, $E(b)$ is true, while $[\forall x]E(x)$ will be true exactly when for every name b whatsoever, $E(b)$ is true.

4. The appropriate truth-functional compounds of the members of the history

subject to the restriction

5. If for two names b_1,b_2, $b_1=b_2$ is in the history, then $E(b_1)$ is in the history if and only if $E(b_2)$ is in the history.[7]

A *world* is any non-empty set of histories.

Now, a world structure can be defined:

A *world structure* contains exactly

(1) One world w.

(2) One of the histories h contained in w (called the "designated" history).

World structures differ from one another according to which world appears, and which of its histories appears.

The world contained in a world structure is meant to correspond to that world which would be the real world if the world structure were the "correct" or "real" world structure (the epistemology of deciding on the correct world structure is not our concern here). The designated history corresponds to what would be the actual state of affairs if the world structure were correct.

But this construction must be further complicated. No constructed entities contain sentences containing definite descriptions, and to define holding for such sentences, we must elaborate the construction. The idea will be to relate such sentences to sentences containing names instead of definite descriptions,—but to which name, if a description is empty? The answer I propose shall in effect say that in such a case the name is arbitrarily stipulated. To make this precise, each history h is paired with each mapping M from the definite descriptions into the names, which results in a collection of histories indexed by mappings (to be called histories*). Two histories* may vary in history, index, or both. And now we modify the previous construction by replacing "history" with "history*." So we find, for example, that two worlds may contain exactly the same histories, but differ in that the histories are indexed differently. But for typographical simplicity hereafter, I shall write "history" and mean "history*."

Now holding can be defined (for sentences without definite descriptions):

Statement	*Holds in a world structure W if and only if*
E, where E ranges over atomic sentences, basic existence statements, and their truth-functional compounds	the designated history contains E
$\sim E$	E does not hold.
$E_1 \& E_2$	both E_1 and E_2 hold.
$E_1 \vee E_2$	either E_1 holds or E_2 holds.
$E_1 \supset E_2$	either $\sim E_1$ holds or E_2 holds.
$[\forall x]E(x)$	for every name b, $E(b)$ holds.
$(\forall x)E(x)$	for every name b, if $(\exists x)(x=b)$ holds, then $E(b)$ holds.
$(\lambda x)(E(x))b$, when b is a name	$E(b)$ holds.
NE	E holds in every world structure W' containing the same world as W (every W' such that $W' \approx W$).
$\Box E$	E holds in every world structure.

Now, a definite description $(\iota x)E(x)$ has a unique referent in a world structure W if and only if for some b, $(\exists y)(y=b) \& (\forall y)(E(y) \equiv y=b)$ holds in W. Let us abbreviate the expression just given by "$RE(b)$," which may be read "b satisfies the Russell condition for 'the E'." Then holding for sentences with definite descriptions is as follows:

Statement	*Holds in W if and only if*
$F \ldots (\iota x)E(x) \ldots$, where F is a predicate of one or more places.	either for some b, $RE(b)$ holds and $F \ldots b \ldots$ holds; or for all b, $RE(b)$ fails, and the mapping which indexes the designated history maps $(\iota x)E(x)$ on to b_1, and $F \ldots b_1 \ldots$ holds.

VIII

Some illustrations and explanations will be in order. Let p be an atomic sentence or truth-functional compound of atomic sentences. Then $\Box p$ holds in a world structure W just if p is in every history whatsoever. $\mathcal{N}p$, however, holds in W just if p is in every history contained in the world in W, for that is the effect for the sort of p hypothesized of canvassing the world structures having the same world as W.

---

[7] This clause serves to give the statements in the basic identity set the status of meaning postulates. Other constraints on histories or worlds would have similar effects.

Plainly, then, the simplest laws will have as at least part of their analysis statements of the form of $[\forall x]\mathcal{N}(Fx \supset Gx)$.[8] Assume for simplicity that F and G are primitive predicates. Then this formula holds in W just if for every name b, referring or not in W, $Fb \supset Gb$ is contained in every history of the world contained in W. $\mathcal{N}[\forall x](Fx \supset Gx)$ easily can be seen to hold on exactly the same conditions. In contrast, $(\forall x)\mathcal{N}(Fx \supset Gx)$ holds in W just if for every name b *which refers in* W, $Fb \supset Gb$ holds in every W' such that $W' \approx W$.

For similar F and G we find that $\mathcal{N}(F(\iota x)Gx)$ holds just if $F(\iota x)Gx$ holds in every W' such that $W' \approx W$. If we let b^{E}_{W} stand for the name assigned to any $(\iota x)E(x)$ in W, either for the reason that $RE\ (b^{E}_{W})$ holds in W, or that $RE(b)$ fails for every b while the indexing maps $(\iota x)E(x)$ onto b^{E}_{W} we can say that $\mathcal{N}(F(\iota x)Gx)$ holds in W just if $Fb^{G}_{W'}$ holds in each W' such that $W' \approx W$.

Consider, however, $(\lambda x)(\mathcal{N}Fx)(\iota x)Gx$. This holds in W just if $\mathcal{N}Fb^{G}_{W}$ holds; that is, just if Fb^{G}_{W} holds for each W' such that $W' \approx W$. Since we need not have $Fb^{G}_{W'} = Fb^{G}_{W}$ and since analogous remarks can be made using $\Box$ rather than $\mathcal{N}$ we see that

(1) $M[E_1((\iota x)E_2(x))] \models (\lambda x)\,(ME_1(x))(\iota x)\,E_2(x)$, where M is either $\mathcal{N}$ or $\Box$

fails, as does its converse. In general, a definite description cannot move in or out of a modal context. The same is not true for names. We have

(2) $M[E_1(b)] \models (\lambda x)\,(ME_1(x))b$, and conversely, where M is either $\mathcal{N}$ or $\Box$.

This behavior points up the fact that since names themselves provide the trans-world-structure criterion of identity, names have "referential constancy," but that descriptions do not have any such referential constancy. For example, it may be that in W_1 $RE(b_1)$ holds, in W_2 $RE(b_2)$ holds, while $b_1 = b_2$ fails in both. Moreover, the definition of holding for definite descriptions makes a sharp distinction between descriptions with a predicate applied to them, and those otherwise occurring in sentences. In the former case, the sentence is "about" the "referent" of the description according to the designated history of *the* world structure for which holding is in question. In the latter case, the sentence is "about" the "referents" of the description according to the designated histories of *a number* of world structures related to the given world structure.

In this way *de dicto* and *de re* modalities are crucially distinct for descriptions, and this produces an effect on the substitution principles valid under identity. Thus as already shown by (1), the following fails:

(3) $t_1 = t_2, E(t_1) \models E(t_2)$, where t_1, t_2 may be names or definite descriptions.

But we do have

(4) $b_1 = b_2, E(b_1) \models E(b_2)$, where b_1, b_2 are names

(5) $t_1 = t_2, F \ldots t_1 \ldots \models F \ldots t_2 \ldots$, where F is a predicate, and t_1, t_2 may be names or definite descriptions.

We find, then, the following arguments invalid:

(6) $\Box(9 > 7)$; $9 =$ the number of planets; therefore, $\Box$(the number of planets > 7),

(7) $\mathcal{N}$ (the man who drank this glass of strychnine was poisoned); the man who drank this glass of strychnine $=$ Jones; therefore, $\mathcal{N}$ (Jones was poisoned),

while this argument is valid:

(8) $(\lambda x)(\Box(x > 7))9$, $9 =$ the number of planets; therefore, $(\lambda x)(\Box(x > 7))$ the number of planets.

This argument may be read: Nine has the property of being necessarily greater than Seven; the number of planets is Nine; therefore, the number of planets has the property of being necessarily greater than Seven.

IX

The question of the relation of the present modal systems to the familiar ones in the literature naturally arises. Results have been obtained for the two modal propositional systems generated by this semantics by my colleague Robert F. Barnes. Exploiting Kripke's semantic characterizations, Barnes has shown that the two modal propositional logics each have the structure of S5. They are related by the expected

---

[8] There are some complexities in the analysis of laws which must await separate treatment.

(9) $\Box E \models \mathcal{N}E$, but not conversely.

Furthermore, the logical modalities "overpower" the nomic ones in the sense that $\Box \mathcal{N}E$, $\mathcal{N}\Box E$, $\Box PE$, $P\Box E$ are all equivalent to $\Box E$; while $\Diamond \mathcal{N}E$, $\mathcal{N}\Diamond E$, $\Diamond PE$, $P\Diamond E$ are all equivalent to $\Diamond E$.

In quantified modal logic, there has been controversy over the validity of the Barcan formula:

(10) $(\forall x)\Box E(x) \supset \Box(\forall x)E(x)$

From the present point of view, neither (10) nor its converse are valid, and the reason is precisely because the system acknowledges merely possible as well as actual individuals. Indeed, we have

(11) $(\forall x)E(x)$ equivalent $[\forall x]((\exists y)(y=x) \supset E(x))$,

so that (10) becomes

(12) $[\forall x][(\exists y)(y=x) \supset \Box E(x)] \supset \Box[\forall x][(\exists y)(y=x) \supset E(x)]$.

The antecedent would hold in a world structure W where the designated history represents an empty domain, i.e., where for every name b, $\sim(\exists y)(y=b)$ is contained in the designated history; but the consequent fails in W if there is some *other* world structure where for some b, $(\exists y)(y=b)$ holds and $E(b)$ fails.

Exactly similar remarks apply if $\Box$ is replaced in (10) with N. However, we *do* have the Barcan formula and its converse for the "square" quantifier with either necessity. Naturally, if we insist that all names refer in all world structures, the square and round quantifiers coalesce, and then (10) and its converse do hold. But then we would also have, for each name b, the valid statement

(13) $\Box(\exists x)(x=b)$,

an explicit acknowledgment of what for classical logic is only shown, not said.

X

A world, except for the one containing all histories, draws the limits of nomic possibility narrower than logical possibility. The histories of the (supposedly) correct world reconstruct the notion of nomically possible alternatives to the actual state of affairs, the latter being represented by the designated history.

This last statement must be qualified, however. Relative to a world structure W, every statement has a truth-value in the sense that it holds or fails in W. If we take the position, though, that statements which contain non-referring names or descriptions may objectively (i.e., in relation to reality) have no truth-value, then no *single* world structure can correspond to the actual state of affairs. A number of them must correspond disjunctively to reality, agreeing on the holding of those statements which are taken to have definite truth-values, but differing on the holding of the others in all possible ways. These world structures will not all have the same designated history; but their designated histories will agree on just those basic level sentences taken to have definite truth values. The world structures will also agree in that they each validate exactly the true laws of nature. Demanding this nomic agreement reflects the commitment that nomic statements are part of the full description of contingent reality. The combined histories of all the worlds of these world structures are the alternative histories to the actual one—or "disjunction" of ones, rather.

This rough picture of the correlation of the constructed entities with reality allows us here to get some picture of the structure of hypothetical reasoning (though far from a full account). In particular, let's consider reasoning with non-modal suppositions cast in the subjunctive; that is, suppositions which we allow may be false.

Such suppositions do not all play the same role. If the supposition clashes with no accepted law, then at most it is counterfactual, but not counternomic. Our intention in making this sort of supposition is to explore alternative possibilities within the limits of the accepted nomic truths. That is, we shall be dealing with the histories of the world which is contained in the correct world structure (assuming for simplicity that only one world structure corresponds to reality), not with all histories whatsoever. The implications of the supposition (aside from trivial ones) are just those sentences which hold in all those histories which contain the supposition (cast in the indicative, of course).

But if, as often is so, we wish to consider the supposition against a background of (some of) what is known to be so, complexities arise. We must face the possibility that the supposition and the known taken together are counternomic. If our intention is still to explore the alternative histories, then we must choose among a number of distinct combinations of supposition with part of what is known, each combination being nomically possible. In this way, the implications of a supposition are controlled by the accepted laws. This establishment of restrictions on hypothetical reasoning by law statements is analogous to the restriction establishing role of logical truths (if we counterfactually suppose A, knowing $\sim A$ to be true, we do not immediately infer to a contradiction).

If our supposition is counternomic, or we accept a counternomic combination of supposition and known, then we have two options, depending on our intentions. On the one hand, we may be engaged in a *reductio ad absurdum* proof, where our inferences are conditioned by the accepted nomic statements as well as by the logical truths. For example, take the following reasoning:

> This copper wire is not passing current.
> Suppose it be (also) under an impressed voltage.
> Then it would be passing current.
> Therefore, it is not under an impressed voltage.

On the other hand, our intention may not be to explore the alternative histories of the "correct" world structure, but rather various sets of world structures, where the world structures in each set validate the same laws and answer to the supposition. That is, so to speak, we would be exploring "science-fiction" worlds.

The behavior of laws just outlined lends itself to the following summary description. The conceptual framework adumbrated provides for two levels of contingent fact, the lower level corresponding to what is true according to which history is designated, the upper level corresponding to what is true according to which world is taken as correctly representing reality. When our intention is to explore the lower level, then the upper level facts (that is, the accepted contingent nomic statements) are deployed in relation to the lower level facts in a manner completely analogous to the deployment of *logical* truths. They are not fully on a par with logical truths, of course, as shown by the second option mentioned in the previous paragraph. The accepted nomic truths can be excluded as ingredients in a reasoning when our intention is of a certain sort. But in the other cases mentioned, the nomic truths enter in and condition the conclusions drawn. This cannot be adequately explained by claiming that reasonings of this sort are enthymematic, with the nomic truths as suppressed premisses, for it is part of the ground rules of these sorts of reasonings, necessary for them to fulfil our purposes, that the nomic truths participate in them. This, as I see it, is the core of truth in the view which takes laws as being, or sanctioning, rules of inference.

XI

We are also able, due to the presence of the square quantifier in our modal system, to state a criterion for identity of intension ("synonymy") for predicates which is not stateable in other systems. Predicates synonymous by this criterion may not be intersubstitutable in "opaque" *epistemic* contexts, such as (taking "believes" in a suitable sense):

(14) John believes that Socrates' nose was pug.

One cannot substitute "retrousse" for "pug" in (14), and obtain a sentence with the same meaning, let alone truth-value, as (14). But this in no way shows that "retrousse" and "pug" are not synonymous. Synonymous terms need be intersubstitutable with a preservation of the meaning of the whole only where the meaning of the whole depends on the meaning of the substituents, but not on other features of them, such as their shape or identity. But the latter would seem to be the case with (14) taken in the opaque sense. Whatever its ultimate analysis, its meaning seems to depend upon the very terms appearing after "that" (and not merely on their content) just as much as the meaning of

(15) John believes the statement, "Socrates' nose was pug" depends on the very terms inside the quotes.

It does seem, then, that intersubstitutability in *all* contexts is not a suitable criterion for identity of intension. Given that, the criterion to be offered presents itself as reasonable.

The criterion is this: Two predicates F_1 and F_2 are synonymous if and only if

(16) $\Box[\forall x](F_1x \equiv F_2x)$, or equivalently
$[\forall x]\Box(F_1x \equiv F_2x)$.

What (16) says is that the extensions of F_1 and F_2 among the possibles as well as the actuals (their "wide extension") are identical, and necessarily so. *De facto* identity of the wide extensions of F_1 and F_2 is not sufficient. It is true, for instance (by standard fictional account), that everything identical with Hamlet is a killer of Polonius, and conversely; but these predicates are hardly synonymous.

Nor is necessary identity of extension among the actuals ("narrow extension"), expressed by

(17) $\Box(\forall x)(F_1x \equiv F_2x)$,

a sufficient condition. First, (17) is not equivalent to (16), for let F_1 be

(18) $(\lambda z)(\exists y)(y=z)$

and let F_2 be

(19) $(\lambda z)(Fz \vee \sim Fz)$.

Then (17) is true, for it is equivalent to

(20) $\Box[\forall x]((\exists y)(y=x) \supset [(\lambda z)(\exists y)(y=z)x \equiv (\lambda z)(Fz \vee \sim Fz)x])$.

But (16) is then false, so that F_1 cannot be substituted for F_2 in (16) with preservation of truth, and surely this means the predicates are not synonymous.

Finally, consider

(21) $(\forall x)\Box(F_1x \equiv F_2x)$.

This, too, is inadequate. Let F_1 be

(22) $(\lambda y)(y=b_1)$

and F_2 be

(23) $(\lambda y)(Fy \vee \sim Fy)$.

Then (21) amounts to

(24) $[\forall x]((\exists y)(y=b_1) \supset \Box(x=b_1 \equiv Fx \vee \sim Fx))$

which would be true in case b_1 alone existed. But still

(25) $[\forall x]\Box(x=b_1 \equiv Fx \vee \sim Fx)$

would be false, and again F_1 could not substitute for F_2 in (16). There are therefore two unacceptable results of the criterion proposed in (21). First, it makes synonymy vary from possible situation to possible situation, depending, for example as just shown, on what happens to exist. Secondly, even in those cases where F_1 and F_2 are "synonymous" by this criterion, F_1 and F_2 may not be intersubstitutable *salva veritate*.

These remarks lead to reflection on the proper form of meaning postulates. If meaning postulates really are to do the job of specifying certain *intensional* relations, then they must employ logical necessity and the square quantifier. Thus, a postulate specifying that R is a symmetric relation is neither of the form

(26) $(\forall x)(\forall y)(Rxy \equiv Ryx)$,

nor any combination of logical necessity with this, but rather of the form

(27) $\Box[\forall x][\forall y](Rxy \equiv Ryx)$, or equivalently, $[\forall x][\forall y]\Box(Rxy \equiv Ryx)$.

Of course, in a language which cannot express modality and the square quantifier, (27) cannot be formulated. In that language we will use (26), which is an implication of (27). But we should not confuse (26) with what we mean to specify by a meaning postulate.

XII

I hope enough has been said to demonstrate the fruitfulness of this approach. Detailed applications of the resources of this system to the detailed analysis of a number of concepts—chiefly, laws, conditionals, and hypothetical reasoning—go too far beyond the reasonable confines of a single article, and so remain to be reported.[9]

Lehigh University

Received December 9, 1970

---

[9] I wish to thank my colleague Professor Robert F. Barnes for much helpful criticism.

American Philosophical Quarterly
Volume 9, Number 1, January 1972

V. THE ONTOLOGICAL STATUS OF HISTORY

NATHAN ROTENSTREICH

I

UNDERLYING the various philosophical approaches to the problem of the nature and facticity of history is the generally accepted assumption that the historical domain is the province or sum total of human events, creations, and institutional crystallizations. Another way of putting this assumption is to say that the uninterrupted process of creation, which is the other face of human events, is at once the content, characteristic, and ground of existence of history. Self-impelled and self-sustained, the historical process is distinguished from natural processes as a novel or superimposed mode of existence. The character of history as a sphere superimposed upon the given realm of nature renders its existence problematical, raising the question: What is the *raison d'être* of the historical process?

A critical analysis of the answers given to this question by Hegel, Marx, and Heidegger is the purpose of the first part of the present study; it leads, in the second part, to the systematic conclusion that the very question concerning the *raison d'être* of history is invalid. This part proposes a possibly more adequate formulation of the problem.

II

The premiss of Hegel's approach to the problem is his conception of history as emergent knowledge, i.e., as the external manifestation of Spirit in time.[1] This notion implies that the existence of history depends upon the existence of a spiritual essence which attains self-knowledge by means of the historical process. The self-knowledge of Spirit is not immediately given. To know itself, Spirit—albeit the absolute real essence—must be unfolded and made externally manifest in process. Essential to Hegel's explanation of the nature and existence of history is the conception of time as the framework in which Spirit attains externality, and through externality, self-knowledge. Hegel's definition of history is at the same time an explanation of its *raison d'être*. From the definition of history as the external manifestation of Spirit in time it follows that, were it not for the existence of history, Spirit would neither know itself nor prevail in the actual world—that is, in the realm of becoming. Although it finds reflection in history, the existence of Spirit is not merely historical. For the Spirit exists as an absolute essence in itself. It is destined to exist as this essence, once its content is made manifest in history and once its self-knowledge is actually adequate to its content. In history, Spirit is diffused, transgressing its internal implicitness, to become determined in the process in time.

What is suggested by Hegel is that history occupies a functional position, a position of mediation, subservient to the attainment of self-knowledge by Spirit. The existence of history *qua* occurrence is useful—useful for an essence which, while using history, at the same time is not historical: its relation to time is not internal but external. Externality in time is for the sake of the internality of Spirit. It is true that Hegel allots a higher status to the externality of Spirit than to the externality of nature. But the point is that neither manifestation of externality is ever allotted an independent status, and that both are subservient to internality. So long as there is mediated becoming for the sake of self-knowledge, history endures. The existence of history is historical.

This conclusion is confirmed by Hegel's statement that the concept of Spirit is the foundation of history and that history is the process of Spirit. This process consists in the transcendence by Spirit of its first self-consciousness, i.e. of that latent consciousness—for the sake of attaining free self-consciousness. The categorical imperative of Spirit, "Know Thyself," is fulfilled.[2] Plainly the implication is that history serves as a field for the fulfillment of a philosophical imperative formulated as the Socratic call. It would thus be in the spirit of Hegel's teaching to paraphrase Fichte's saying about nature as the matter for the fulfillment of duty, and say that history is the field for the fulfillment of the imperative "Know Thyself," by Spirit.

---

[1] G. W. F. Hegel, *Phänomenologie des Geistes*, ed. by G. Lasson (Leipzig, 1921), p. 520.

[2] See the *Vorlesungen über die Geschichte der Philosophie*, in Glockner's edition of Hegel's, "Sämtliche Werke" (Stuttgart, 1928), Vol. XIX/3, pp. 19, 103, 104.

Clearly we approach here the basic antinomy or ambiguity in Hegel with reference to the relation between time and essence. See the present author, "The Epochal and the Essential Aspects of Philosophy," *The Review of Metaphysics*, vol. 23 (1970), pp. 699–716.

Our criticism of this view of Hegel's will not proceed along the lines of Marx's criticism of it, for reasons which will be apparent presently. Our criticism is not designed to put forward an intra-historical factor—in Marx's case, economic production—in place of an extra-historical factor—in Hegel's case, Spirit—as the *raison d'être* of history. It is designed to show that this theory is untenable for the following reasons:

(*a*) Assuming the existence of a real essence called Spirit, there is no ground for concluding that just history, *qua* process in time, is demanded by the nature of Spirit as the mediating sphere subservient to its attainment of self-knowledge. Granting that self-knowledge is an essential property of Spirit—this does not imply that only in the process of temporal succession, i.e., only in history can the actualization or emergence from potentiality into actuality of self-knowledge be achieved.

(*b*) The assumption that consciousness and self-consciousness are properties of Spirit is not incompatible with the Aristotelian assumption that there exists a pure activity (*actus purus*) in which there is no potentiality whatever. It may be maintained that whatever is a potential property of Spirit is actual. Thus Hegel's characterization of Spirit is compatible with a mode of self-reflection which does not entail the external manifestation of Spirit in time. From free self-knowledge as the property of Spirit, history as a means of acquiring that property does not follow of necessity.

(*c*) This being so, there is no justification for maintaining that whereas Spirit is creative *ab initio*, history is merely ancillary or functional. The most Hegel can claim is this: Given the existence of process in time as a primary datum, it follows that the finite human beings who subsist in the temporal process cannot comprehend the nature of Spirit in a pure complete act. There being no immediate, and, as such, complete contact between men and Spirit, men discover Spirit in the medium of their existence, i.e., in time. Because man subsists in time, and because he must therefore unfold Spirit so as to grasp it, Spirit must be unfolded in time. But it is not in terms of this contact between Spirit and men that Hegel approaches the problem of history. To him, man is but the actualizer of Spirit. And his ascription of an instrumental or ancillary status not only to history but also to man results in what might be called a hypostasis of Spirit.

(*d*) Hegel does not answer the question: What brings the historical sphere into being? The only question to which Hegel's approach affords an answer is: What are the stages through which the process of man's confrontation with Spirit must pass? or, How does mankind gradually attain self-knowledge? A *post factum* characterization of the connection between spiritual contents and history cannot disclose the reason for the very being of history.

III

As contrasted with Hegel's approach to the problem of history and as a conscious attempt to correct its flaws, Marx's approach may be said to have developed from an intra-historical starting-point. For it is true that Marx does not treat history as the handmaid of an extra-historical essence. But Marx's starting-point turns out to be not quite as intra-historical as it seems, at least not if "intra-historical" is taken in the strict sense.

Even the fact that in his early writings Marx names his system both "historical materialism" and "practical materialism" suggests that for him history is connected with practice, or "revolutionary *Praxis*," as he calls it. If Marx traces the origin of history to the causal agency of an intra-historical act, it is because that act is executed in history and because its execution presupposes diverse historical circumstances as its precondition. But there is reason for maintaining that this act can be called "intra-historical" only in a qualified sense.

An immanent approach to the problem of history would seem to imply the twofold assumption that (*a*) the historical process is the sum total of acts by which it is generated, and that (*b*) from the point of view of the process as a whole, no particular historical act occupies a unique status. But Marx's treatment of the problem is not controlled by this assumption; for he refers not only to acts within history but also to an act, or to the structure of an act, by which history itself is created. It is in these terms that he discusses the status of economic production. The substance of his teaching is summarized in two key passages. The first maintains that men can be distinguished from animals by consciousness, religion, or anything we choose; yet men themselves first begin to distinguish themselves from the animals when they begin to *produce* their means of life. In producing their means of life, men indirectly produce their material life itself.[3] The second passage formulates the conclusion implied in the first, namely that the first historical act is the production of means of satisfying the necessaries of life, the production of material life itself, and that this is a basic condition of all history.[4] Marx, too, investigated the "why" and "wherefore" of the existence of history, and his investigation led him to "the first historical act." The status of history is instrumental. History exists for the sake of production, designed, directly, to satisfy man's wants and, indirectly, the source of new wants. Were history not to be created, the wants of man would not be satisfied. Its creation is demanded by man's very structure, his physical or natural structure. Given his conception of the origin of history in an initial act which produces means of satisfying human wants, one can understand why, in his doctrine, Marx endows economic processes with causal efficacy in the on-going historical process. He regards the historical-economic processes connected with the means of production as concrete realizations of the first historical act, i.e., of that act causing the creation of the very historical domain. Because he does not posit the dependence of the historical process upon a hidden spiritual treasure which it reveals, Marx no longer treats the process in terms of the emergence from potentiality into actuality of a latent content. Yet he does imply that the generation of history has a final cause as well as an efficient cause. Satisfaction of the wants of man is the final cause of the historical process, and production is its efficient cause. The efficient causality of production is entailed by the final causality of want-satisfaction. To treat history in these terms is to imply that it occupies an instrumental status. In Marx's system, however, the instrumentality of history does not imply that history itself will come to an end once its task of making Spirit manifest and facilitating the fulfillment of self-knowledge is accomplished. As enduring as its twofold cause is inexhaustible, history will last as long as men seek satisfaction of their wants and produce the means of satisfying them.

The major flaws in Marx's treatment of the problem of history seem to be the following:

(*a*) Marx fails to offer an account of the historical complex as related backward to the past as well as forward to the future. With the aid of Marx's correlation of history with "*Praxis*" the most one can explain is men's active orientation to, and planned construction of, the future. What cannot be explained on Marx's premisses is man's contemplative orientation to, and dependence upon, history in the sense of the flow of events from the past to the present.

The hypothesis of a "first historical act" is no solution to the problem presented by man's historical consciousness, that is, by his conscious dependence on, and directedness toward, the process thus far. The problem of man's historical consciousness cannot be solved by pointing to an act of production oriented to the future. That historical consciousness or orientation to the past is derivable from practical orientation to the future, and that, as such, it occupies a secondary status, is the consequence of Marx's line of reasoning. Yet the derivation of historical research from economic processes is by no means necessary. Even granting that the task of historical research is to discover the dependence of historical processes upon economic processes, the execution of this task does not entail the conditioning of research *qua* research by its subject-matter, i.e., economic process.

(*b*) There is no justification for Marx's manner of isolating from the historical complex as a whole a single element of content—the productive act—and of assigning that act a unique causal status in terms of its content. In itself, Marx's assumption of human wants which demand satisfaction as the final cause of history, and of a process of economic production subservient to the satisfaction of those wants as its efficient cause, this twofold assumption implies the existence of a concrete nexus of which no single constituent factor may be elevated to the status of its ultimate ground. This concrete nexus includes as one of its inherent factors the verbal communication which constitutes men's means of informing one another of their wants and of planning the process of production with a view to satisfying those wants. Neither the act of giving information, nor the act of understanding the information given, nor the ability to plan production in conformity with that understanding, can be accounted for by the fact that men have wants and generate a process of production. On the contrary, it is the process of production which can be accounted for by man's awareness of his wants. And though, or precisely because, it may be oriented to the adjustment of production to wants, the position of understanding is always partially independent of the nexus of which it is a constituent and in which it plays a regulative role. It is to understanding that the process of production owes its instrumental status. Adjustment of product to want presupposes man's ability to distinguish between the instrumental function of a hammer *qua* product and the instrumental function of a loaf of bread *qua* product, to take an elementary example. Were there no understanding of the difference between the functions of diverse products in relation to the diverse wants they are designed to satisfy, there would be no process of production or, at least, no adjustment of its efficient cause (productive process) to its final cause (want-satisfaction). The regulative position of understanding is irreducible to either one or both of the two kinds of causality in terms of which Marx treats the problem of history.

---

[3] Karl Marx and Friedrich Engels, *Die Deutsche Ideologie*, in V. Adoratsky's *Historisch-Kritische Gesamtausgabe* (Berlin, 1932), vol. I/5, p. 10.

[4] *Ibid.*, p. 17.

(*c*) That the irreducible status of understanding which Marx overlooks follows from his own assumptions can also be shown from another viewpoint. On the basis of his own recognition of the non-biological nature of the process of production, Marx should have posited conscious or reflective mediation as the necessary ground or precondition of production. Marx's assumption of the primacy of production vis-à-vis history necessarily entails the conclusion that the producer is man as endowed with understanding of his wants. In the act of production, man directs himself toward a thing which still has not come into existence. He does this in the expectation that by making the absent product present, he will satisfy the want whose satisfaction it is the function of that product to fulfill. In virtue of his expectant or anticipatory directedness, through production, to a non-present product, man is separated in status as well as in actual time from his object. For the producer exists in the present in a state of preparation for the product to be produced, whereas the product will come into existence only at some time distant from the present. The gap between producer and product, subject and object, is more profound than the gulf between the wants of man and the products whereby they are satisfied. To fulfill its function, the process of production presupposes the activity of consciousness. Wilhelm von Humboldt characterized the activity of consciousness as follows: To engage in reflection, the mind must momentarily arrest its incessant activity, that it may grasp its impression into a unity and, thereby, set that unity over and against itself as an object.[5]

Production remains an object of intentionality even while fulfilling an instrumental function in relation to the want which causes its generation. In this respect consciousness—which never occupies an instrumental status only in relation to its object—constitutes the necessary precondition of production. But because consciousness—of men who exist in history—is an intra-historical factor, neither consciousness nor its acts of understanding can be conceived as "the first historical act," and are rather to be regarded as the precondition accompanying all historical acts. If, however, all historical acts are accompanied by the activity of consciousness or understanding as a precondition from which they are inseparable, then the hypothesis of a "first" historical act does not seem subject to demonstration. It is more probable that all historical acts take place while men are already in history. But what is important to remark is that if historical acts are to be assigned an intermediary status, it is because they are accompanied and conditioned by understanding consciousness. Why consciousness is the *conditio sine qua non* both of historical activity oriented to the future and of historical contemplation oriented to the past is a question which will be treated below.[6]

IV

Time, in Heidegger, should be regarded as the horizon of every understanding and interpretation of Being.[7] This implies that the starting point and the foundation of Heidegger's approach to history is time. Heidegger reduces history, from the outset, to a mode of relatedness to time. Still, in his discussion of historicity, he makes the statement that Being is historical and presents this statement as a fundamental ontological assertion. Formulating his correlation of temporality with historicity, more specifically as the finitude of temporality, Heidegger posits this finitude as the implicit reason for the historicity of Being.[8] By finitude he means Being on the verge of death.[9] The term "finitude" denotes not the objective fact that the duration of man's existence is finite, but the consciousness of finitude conceived as the constant anticipation of man's end or doom. Considering the status it allots the future as the dimension of doom, Heidegger's system might, perhaps, be characterized as "negative eschatology."

---

[5] See "Über Denken und Sprechen," in Leitzmann's edition of *Wilhelm von Humboldt's Werke* (Berlin, 1908), Vol. VII/2, p 581.

[6] Compare the present author's book, *On the Human Subject, Studies in the Phenomenology of Ethics and Politics* (Springfield, Ill., 1965), pp. 30–51.

[7] M. Heidegger, *Sein und Zeit* (Halle a.d.S., 1927), p. 17. Reference is made only to this book and not to the subsequent statements by Heidegger.

The primary manifestation of primordial and real temporality is the future. Or, else, the future constitutes that coming in which existence comes to the possibility of its proper, most concrete, mode of Being. Though he uses the term "possibility" in connection with the future, Heidegger clearly does not present the future as a dimension in which the active possibilities latent in existence find realization, or emerge from potentiality into actuality. What he means by the term "possibility" is not a latent potentiality to be made actually manifest, but an immanent reality, the most primary reality (death) to be prepared for by way of anxiety. The true content of Being may be said to lie in its end or doom rather than in man's active standpoint which calls into being something hitherto nonexistent. So central is the future's role, that it alone arouses the present. As for the present, it is no more than the point at which, or the dimension in which, the determination of existence "for the sake of" something is accomplished.[10]

In Heidegger's system the concept of historicity is bound up with the concept of individual existence. Were he not to assume this connection, he would have no justification for elevating the anticipation of the end, which is the peculiar property of individual consciousness, to so central, nay crucial, a status. Further insight into Heidegger's handling of the problem of history is afforded by his reference to historicity in the "un-true," "un-real," sense of the concept, by which he apparently means the translation of personal doom and consciousness thereof into what he calls the extension of destiny.[11] What the latter phrase probably alludes to is an impersonal power to be contradistinguished from death. For death is not an impersonal power opposed to personal existence, but a latent possibility immanent in personal existence itself. Only the personal is primary. Whatever transcends the personal is secondary, be the personal transcended toward the infinity of physical time or be it transcended toward historicity in the untrue, unreal sense of destiny, circumstances, indeterminate forces, etc.

(*a*) To judge from Heidegger's powerful influence over our generation, one might get the impression that he deserves credit for having turned the spotlight on man's historicity and for having initiated the systematic quest for an explanation of history's status. There is, to be sure, good reason for crediting Heidegger with what might be called an internalization of history, unprecedented in Hegel and Marx. In his system, history is treated neither in terms of a spiritual essence which it makes externally manifest, nor in terms of human wants conceived as its "prehistoric" substratum. Only with history is Heidegger conconcerned; with history from the sole viewpoint of human existence in time. And, closing his focus still further, Heidegger treats human history in such a way as to give unequivocal priority to man's relatedness to the future. But the question is whether Heidegger's concept of historicity is clear. That the answer is *no*, can be shown by considering his attempts to define what he means by historicity. To begin with, he invokes the familiar distinction between history as a process of events and history as the study or science of that process, and argues that the problem of history pertains to the former, not to the latter. He then adds that to understand how history *qua* process or occurrence becomes a possible object of historical research, one need only examine the nature of the process.[12] So far, so clear. That is, one can understand the view which claims that historical occurrence must be distinguished from the study thereof. One can even understand the inclination to assert the primacy of history *qua* occurrence on the grounds that existence (historical events) is prior to the science thereof (the study of historical events).

---

[8] *Ibid.*, p. 386.
[9] *Ibid.*, pp. 327 ff.
[10] *Ibid.*, pp. 346 ff.
[11] *Ibid.*, pp. 387 ff.
[12] *Ibid.*, pp. 392 ff.

But it is more difficult to understand what bearing this distinction can possibly have on the meaning Heidegger assigns the concept of historicity. What connection can there be between Heidegger's concept of historicity and the familiar concept of history as that process of human events which constitutes the object of historical science? The latter connotation of history would seem to correspond precisely to what Heidegger terms history in the "untrue," "unreal" or impersonal sense. Heidegger is concerned with man's historical consciousness as anxious anticipation of the future as the dimension of doom. Whatever does not come under this heading, i.e., whatever transcends the personal, is not historical in the "true" sense. Yet the proper sphere of history is precisely the province of public events, creations, and institutions. And as a public domain, history surely includes impersonal elements. Thus in the realm of history one finds such institutional crystallizations as the State and the Law, whose very nature entails impersonality and whose sphere transcends the confines of the individual man's future relatedness to himself. The same would apply to language, that constituent of historical reality upon which Heidegger so often dwells and dilates. Crucial though its part in personal realm is, language is clearly also an impersonal sphere. Conditioned by impersonal history in general and by the impersonal history of language in particular, personal verbal expression is a particular focusing, crystallizing interpretation of an impersonal medium. Nor will it do to explain the impersonal away as a sphere in which manifold personal lines converge. For the point exhibited by language is precisely that the impersonal is logically and factually prior to the personal. Because the creation of the personal presupposes the impersonal as its background and medium, the impersonal cannot be conceived as a sum of personals.

Contrary to the implications of Heidegger's appeal, his system turns out to be incapable of accounting for the historical domain *par excellence*, i.e., for the realm of history in the systematic as well as in the current sense of the term. To proceed on the assumption that every objectivization is a distortion, depravation, and renunciation of authenticity is to turn one's back on history altogether. By Heidegger's romantic treatment, the problem of history is shelved, not solved. Contrary to what its program leads us to expect, Heidegger's system fails to confront, let alone to solve, the problem Hegel and Marx had set out to explore.

(*b*) If Heidegger's concern might be viewed as irrelevant to the problem at hand, even waiving this objection, one would still have to question the validity of Heidegger's statement that time is the horizon of every interpretation of Being, or the principal realm of awareness from which we set out to understand Being. For one thing, the meaning of this saying is by no means clear. Does it imply that temporal occurrence and succession constitute the background of our interpretative acts because these acts, being actual, are necessarily executed in time? Such an implication would almost be a truism, since no one would claim that the occurrence of an act does not occur in time. But perhaps Heidegger's expression, "horizon of every interpretation," means to suggest that time is the content or theme, and not only the background or framework, of every interpretative act. This would imply that time *qua* content is the mother of interpretation, and that the most essential dimension of time, viz., the future, is the revelation of the totality of Being. This might not be a truism; but neither is it true.

Interpretation is one mode of reflective intentionality. All acts of reflective intentionality occur against the background of time, and some have time as their object. If time is the content or theme toward which reflection directs itself, then it occupies the status of an object over and against reflective intentionality, and not the status of a content which calls reflection itself into existence. Although his own concept of conscious anticipation of doom presupposes reflective intentionality, Heidegger seems to suggest that consciousness of *doom* is the mother of *consciousness* of doom; he apparently believes that reflective intentionality is essentially consciousness of doom. This seems to be false. Even consciousness of doom is a mode of interpretation. Since every interpretation has an intrinsic structure which is irreducible to the structure of its content, the structure of interpretative directedness toward time is irreducible to the structure of time.

Every interpretation entails an analysis of certain thematic data, as, e.g., Heidegger's interpretation entails an analysis of the future and of the present. In addition, every interpretation involves an attempt to establish relations and connections between the data analyzed. The structure of interpretation is thus characterized by two basic trends, one analytic, the other synthetic; one distinguishing between diverse contents (as, in Heidegger's case, between the content of doom and the content of anticipatory directedness toward doom), the other establishing relations between contents (in Heidegger's case, placing the future in a relation of "primacy" *vis-à-vis* the present, and the present in a relation of "for the sake of" the future). Neither the discerning-analytical trend, nor the relating-synthetic trend are modeled on the pattern of the discerned and related contents. Both are intrinsic trends of reflective consciousness; reflective consciousness is irreducible to consciousness-of-content, be that content as it may. No matter how much weight it carries from an existential point of view, from the viewpoint of its systematic status a *content* of consciousness remains a content. Consequently, time cannot be conceived as the horizon of, i.e., as occupying a higher status than, the consciousness directed toward it. Accordingly, our second objection to Heidegger's treatment of history is twofold: (*a*) his ontological premiss is invalid; (*b*) his historiosophical conclusion is groundless.

(*c*) Here again one can carry the critical analysis still further and show that just as Heidegger misconstrues the status of time as a theme, so he misconstrues the status of its three dimensions. Beyond the obvious differences betweeen past and future, their positions are alike, perhaps even equal, in one central respect which Heidegger fails to take into account. Contrary to the implication of Heidegger's suggestion that whatever has been is, in a sense, derivable from what is to be, there is reason to maintain that the position of the future is prior to the position of the past. The reason is this: Underlying both man's relatedness to the future and his relatedness to the past is a standpoint transcending the present moment, here and now. Both directions of relatedness presuppose transcendence of the present as their precondition. This implies (*a*) that transcendence itself cannot be derived from either direction of relatedness and (*b*) that transcendence in the direction of the past cannot be derived from transcendence in the direction of the future. Given the transcendence of the here and now as the common denominator of both directions of relatedness, history cannot be treated exclusively in terms of either direction. To reason from transcendence of the present, i.e., from reflective consciousness which, recognizing the transience of the present moment, conjures up for itself both past and future, is to preclude the ascription of primacy to any particular temporal dimension, be it past, present, or future, as the key to the meaning of history. By this we do not mean to deny the difference between the dimensions of time. We mean only that this difference lies not in their status, but in their diverse contents on the one hand, and in man's diverse modes of relatedness to their respective contents on the other.

The differences between the contents of past, present, and future entail no differences between their ontological-constitutive positions in relation to history. The grounds of history as a flow of occurrences along the dimensions of time lie not in time itself, but in man's cognitive relation toward its three-dimensional domain and the contents thereof. To understand the nature of history one must go outside time, to man's historical consciousness, that is, to his orientation toward past, present, and future on the one hand, and toward creations crystallized against the background of occurrence in time, on the other.

V

This does not mean to imply that historical consciousness harbors a solution to the problem of history as it is understood by Hegel, Marx, and Heidegger. The alternative approach to the problem we are about to broach involves the abandonment of the very attempt to determine the *raison d'être* of history. The starting point of this approach will be the given fact of man's existence in, and practical or contemplative orientation to, the flow of historical occurrence. Its task is not to explain the origin of history in terms of its final or efficient cause, but to analyze the given historical complex in terms of its underlying assumptions.[13]

Man's relation to time as the form of succession is the foundation of the historical domain. In his day-by-day experience man is confronted with the fact of flux or temporal succession, i.e., with the fact that event '*b*' follows event '*a*'. In addition to the formal element of succession man recognizes an element of content in virtue of which event '*b*' is connected with event '*a*'. The two aspects of man's relatedness to time—recognition of temporal succession and discernment of inter-connection between successive phenomena—are two sides of the same shield. For to circumscribe a phenomenon as later or earlier than another phenomenon, is *ipso facto* to correlate the two phenomena. In perceiving succession one not only confronts the fact of temporal transition, but also takes the first step toward clarifying the relations constitutive of transition. Discernment of the reciprocal relation which obtains between the stages of process is thus implicit in the recognition of process.

---

[13] On the position of time see the present author's *Between Past and Present, an Essay on History* (New Haven, 1958). Consult also Paul Weiss, *History: Written and Lived* (Carbondale, Ill., 1962), pp. 141 ff., 197 ff., 217 ff.

As entailing attentive directedness toward, and discernment of, process, man's twofold relatedness to time would not be possible were he totally immersed in the process. Were our process-discerning relatedness to time merely a matter of process, we could not confront process as the object of our discernment. As a relation of time-discernment, man's relation to time implies a distinction between man—the discerning subject—and time—the discerned object. Man may therefore be said to occupy a double status in relation to the flow of events in time: on the one hand, he lives in, and moves with, the process, but on the other, he observes it and as an observer, he occupies a status of detachment from the observed process. The condition of man's reflective relatedness to time lies not in existential engagement in the object, but in ideal intentionality toward it.

This affords further grounds for the rejection of an attempt to explain history in terms of an extra-historical constitutive factor, be it a spiritual content (Hegel) or a material want (Marx). For the same reason one must also reject the explanation of history in terms of an intra-historical but nevertheless non-historical factor—i.e., the individual's anxious anticipation of his own future (Heidegger). An adequate account of historicity must take its departure from man's basic relatedness to objects in general, and to time *qua* object in particular. The polar structure of this relation indicates that man's ontological position is to be conceived as neither outside, nor on the threshold of, nor totally inside, time. In the light of these considerations, history is a particular interpretation given by man to his directedness toward time.

Understood as man's interpretation of time, history may be defined as (*a*) the sum-total of man-made processes in which such human creations as language, law, economy, state, etc., are crystallized, and as (*b*) the sum total of those processes in which man harnesses such natural factors as rivers, weather, laws of nature, etc., to his own ends. The distinction between these two components of the historical complex may be formulated as follows: Whereas man-made crystallizations such as language, state, etc., owe both their creation and their meaning to human agency alone, natural factors integrated into human reality do not owe their entire meaning to man's integrating activity. Thus, for example, the currents of rivers navigated by man would retain a core of meaning even were man not to harness them to his ends or integrate them into his works. But beyond the difference between them, both aspects of history as the sum total of man's creations share one feature in common, namely—publicity. In this respect, history may be defined as a public domain, i.e., as the province of men's concerted intentionality and effort toward crystallizing creations which transcend the individual domains of their creators. Impersonal publicity is thus one aspect of *content* characterizing the historical complex.

From here it is possible to move to the assumption of the precondition of history. As the province of public creations crystallized beyond the private domains of individual men, history presupposes a creative process which neither begins nor ends in the present, which in turn presupposes that the structure of time is open. From the point of view of its structure, time permits an unimpeded flow of the creative process along either or both of its channels. The open structure of time underlies man's relatedness to creations crystallized over the ages by the effort of generations. In relating himself to the collective creations of mankind, man adopts a position within their framework as well as over and against it. This does not mean to imply that man's position is subject to shaping in the narrow sense of historical determinism. But it does mean to imply that his position is exposed to the regulative control of historical creations. Thus, for example, a measure of thematic determination is clearly involved in the control of man's ideas or orientation by the collective creations such as language or the state to which he relates himself.

The foregoing characterization of the historical domain indicates that history, understood as a particular interpretation of man's relation to time, presupposes as its precondition man's cognitive relatedness to crystallized contents on the one hand, and to the open structure of time on the other. To be sure the question to which this assumption affords an answer is not, "What is the reason for being of history?" but rather, given the existence of history, "What is the condition of its possibility?" The suggested answer is that the precondition of the possibility of history consists in man's reflective relatedness to the collective creations of mankind and to the temporal structure which constitutes the background of the creative process.

This reflective relatedness is partly contemplative, partly practical. There is a practical-active advance from the present to the future, and a contemplative-retrospective concern with the past as the province of completed creations. Beyond the obvious differences between active orientation to the future and contemplative orientation to the past, both faces of the historical complex presuppose the open structure of time whose openness is interpreted now as the future and now as the past. We have already observed that transcendence of the present moment is the precondition of relatedness to past and future alike. We can now add that reflective consciousness confronts the openness of time through its contemplative directedness toward the past and its intervening directedness toward the future. The difference between past and future is accounted for by the difference between the modes of reflective relatedness, now contemplative and now creative. In transcending the present instant toward either future or past, one releases oneself from immersion in time. Reflective detachment from time-present is the condition of reflective relatedness to both time-past and time-future.

VI

The fruitfulness of an approach to the problem of history controlled by the concept of reflective relatedness can be indicated by pointing out some concrete elementary constituents of everyday historical experience which invite interpretation in terms of reflective relatedness to collective creation, on the one hand, and to the open structure of time on the other. As our first example, let us take that sense of process which leads man to recognize the openness of time.

To perceive the passage of generations is to perceive that one stands in the midst of this passage, and concurrently that one stands in the midst of time. Latent in man's awareness of the succession of generations are the seeds of his reflective relatedness to time in the philosophical sense of this concept. Implicit in the very awareness of biological succession is the historical consciousness of the open structure of time. Once you realize that men have lived before your time and will live after your time, you have perceived that the particular point where you stand is neither the beginning, nor the end, but in the middle of time. In other words, you have discerned that the structure of time is open. We are referring to man's realization that "one generation passeth away and another generation cometh." Although the succession of generations is, in a sense, a "pre-historical" or biological datum, man's interpretation of it is historical.

Here, then, is an illustration of an elementary human experience historically interpreted in terms of time's open structure by a creature endowed with reflective consciousness. This is not to deny the obvious fact that, were he not endowed with historical consciousness in the first place, man would not interpret his awareness of the succession of generations in historical terms. Albeit the occasion of historical consciousness, experience of alternating generations is not its cause. But if our example has illustrated the bearing of historical consciousness *qua* relatedness to time as an open structure upon man's elementary experience, then it has served its purpose.

What about the other element of historical consciousness, namely relatedness to the collective creations of mankind? Is there concrete evidence of its significance for man's elementary historical experience? The affirmative answer follows from an analysis of language as a given component of man's concrete reality, and as a component whose nature is historical. Language is a collective, impersonal creation whose existence is given prior to its use by a particular person for the purpose of self-expression or speech. Language is not created, but only released or realized by speech. In this respect, man's relatedness to language involves a confrontation with a collective creation of the many for the sake of the many. Only as one of many can the individual partake of this public creation. Our purpose here is not to analyze the systematic significance of language, but to establish the relation between language as a collective creation and historical consciousness as a mode of reflective relatedness. The existence of language is contingent upon the speaker's transcendence of the act of speech here and now, and upon his directedness toward what lies beyond. Reflective relatedness to the beyond is the condition not only of the existence of language, but also of its meaning as the theme or content of intentionality; and reflective intentionality whose theme is language is an outstanding example of historical consciousness proper. Implicit in man's confrontation with language is an awareness (*a*) that he is not its creator, (*b*) that he has inherited it from past generations, and (*c*) that he is subject to determination by its contents as an historical creation. From this it follows that both phases of historical consciousness—i.e., reflective relatedness to collective creations and reflective relatedness to time's open structure—find their manifestation in man's relatedness to language. Unlike man's awareness of the succession of generations, which derives historical significance from a pre-historical given, man's confrontation with language exposes the essentially historical meaning of the lingual datum.

The foregoing analysis does not afford an answer to the question *why* history exists. Its more modest task, as stated, was to illustrate the bearing of historical consciousness upon concrete experience by indicating given components of human experience which invite historical interpretation.

In view of its bearing upon empirical components of historical reality, reflective relatedness seems to afford a vantage point from which one can survey the full sweep of the historical process. Both faces of history—as the so far completed process from past to present, and as the on-going process from present to future—presuppose reflective relatedness as the condition of their possibility. And each face represents a realization of reflective relatedness. To put it another way, both contemplative directedness toward recorded history, and creative directedness toward the on-going process of historical occurrence presuppose man's relatedness to the open structure of time. As the object of reflective intentionality, time becomes the theoretical and actual condition of history. Thus man's double standpoint in history, as both the denizen and the detached observer of its domain, represents one realization of reflective relatedness. This implies that the twofold condition of history consists in man's contemplative—or reflective—directedness toward the past and creative—reflective just the same—directedness toward the future.

The Hebrew University of Jerusalem

Received August 10, 1970

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 1, January 1972

VI. BASIC ACTIONS AND SIMPLE ACTIONS

JANE R. MARTIN

IT is a commonplace for philosophers concerned with the concept of action to distinguish between two sorts of actions—actions which are done by doing some other action and actions which are done without doing some other action. The latter sort of action they call *basic* or *simple*; the former sort they call *non-basic* or *complex*. There is disagreement at the present time over the proper formulation of this distinction.[1] Rather than discuss the question of formulation here, however, I want to examine the distinction between basic and non-basic actions with a view to casting doubt on a thesis which has been central to it and which is independent of any particular formulation of it. I shall not argue for the abandonment of the distinction for I think it is an important one. I shall, however, argue that the thesis that basic actions are necessarily simple actions—a paradigm case being that of moving a limb—be abandoned. It goes without saying that if this thesis is abandoned the distinction will have to be conceived of rather differently from what it has been.

The thesis with which I shall be concerned is part of what I shall be calling here the "standard view" of basic actions.[2] It is widely accepted that moving a limb or a finger is a standard case of a basic action. Now although it is acknowledged that not even a standard case of a basic action is a basic action in every context (moving my arm, for example, would be a non-basic action if I moved my arm by pushing it up with my other arm), it is generally agreed that there are limits on what can be a basic action. There may be more candidates for basic action than moving a limb, but the number of qualified candidates is definitely restricted. Opening a door by turning the knob, moving a stone by kicking it, signalling by raising one's arm, knotting a tie, dancing the polka, to name just a few, are actions which in no circumstances are basic.

It seems to be generally accepted also that basic actions, or rather the abilities of which such actions are exercises, are not acquired or learned but are "gifts."[3] The actions cited above (moving a stone, knotting a tie, etc.) which are assumed by advocates of the standard view to be non-simple (complex) and which are taken by them to be paradigm cases of non-basic actions are on the view I shall be proposing here capable of being performed as basic actions. These and countless other actions which on the standard view are complex but which on my view can be basic actions are surely not gifts or exercises of gifts. If the present discussion is to proceed, then, it is essential that the gift thesis be disputed. I shall turn to it now and only after I have disposed of it shall I be in a position to argue my case.

I. ARE BASIC ACTIONS GIFTS?

I read a chapter of *Tom Sawyer* to my sons last night. I take it this was an action of mine. Is it something I did by doing something else? Are some other actions included in it or does it consist in other actions? Surely my reading the first page of the chapter is included in my action of reading *Tom Sawyer*, and included in that action is my action of reading the first word on the first page.

---

[1] See Arthur C. Danto, "Basic Actions," *American Philosophical Quarterly*, vol. 2 (1965), pp. 141–148. Reprinted in Alan R. White (ed.), *The Philosophy of Action* (Oxford, 1968), pp. 43–58 and in Myles Brand (ed.), *The Nature of Human Action* (Glenview, Ill, 1970), pp. 255–266. For criticisms of Danto's account see Frederick Stoutland, "Basic Actions and Causality," *The Journal of Philosophy*, vol. 65 (1968), pp. 467–475; Myles Brand, "Danto on Basic Actions," *Nous*, vol. 2 (1968), pp. 187–190 and pp. 224–228 of *The Nature of Human Action*.

[2] The standard view is a composite of the recurrent themes in the literature on basic actions. See especially Danto, *op. cit.*; but also Danto, "What We Can Do," *The Journal of Philosophy*, vol. 60 (1963), pp. 435–445, "Freedom and Forbearance" in Keith Lehrer (ed.), *Freedom and Determinism* (New York, 1966), pp. 45–63 and *Analytical Philosophy of Knowledge* (Cambridge, 1968), *passim*; A. I. Melden, *Free Action* (London, 1961), ch. 4, Richard Taylor, *Action and Purpose* (Englewood Cliffs, N.J., 1966), pp. 117–119. In citing these writers I do *not* mean to suggest that any one of them subscribes to every aspect of the view I am outlining here.

[3] Danto's term; *Analytical Philosophy of Knowledge*, *op. cit.*, p. 67.

Does that action include some other or is it basic? Did I read "the" by doing something else or not? It seems to me that my reading the word "the" was a basic action or if it wasn't my recognizing the pattern made by "the" was. But then, the standard view of basic actions is wrong in suggesting that basic actions are gifts. For as a matter of fact reading "the" is not itself a gift but is, at least in most cases, the result of learning of some sort or other.

But of course proponents of the standard view may not be satisfied with my stopping point in my breakdown or analysis of reading a chapter of *Tom Sawyer*. They might insist that I read "the" by moving my eyes and that this, not reading "the," is my basic action. Yet can they consistently maintain this? My eyes do perhaps move when I read but isn't their moving on a par with my muscles flexing when I raise my arm? The latter are considered by advocates of the standard view to be happenings or goings on, not actions of mine. Aren't the former happenings also? To be sure, I can move my eyes when I want to; that is to say my moving my eyes can be an action of mine. But of course I can flex my muscles when I want to: my moving them can also be an action of mine. Nor will it do to point out that I don't learn to move my eyes, that this is an exercise of a "primitive ability,"[4] hence a basic action. For it is only where something is independently acknowledged to be an action that its being the exercise of a primitive ability and not one acquired through learning can mark it as a basic action. After all, physiological happenings—the very things philosophers of action want to distinguish all actions, hence basic actions, from—are not the result of learning either.

If my *Tom Sawyer* example is unconvincing consider my raising my right eyebrow. I take it this is an action of mine.[5] It is, moreover, something I do without performing some other action. To be sure, certain muscles flex, but as we have seen, muscle flexings are not on the standard view actions of mine. My raising my right eyebrow is, then, a basic action. Yet it surely is not the exercise of a primitive ability, an ability which is a gift and not acquired. I taught or trained myself to raise my right eyebrow: I stood before a mirror, held my left eyebrow down, pushed my right one up, and so on and did this for weeks on end until finally I could just raise my eyebrow without doing anything else.

Now according to the standard view there are no "recipes"[6] for basic actions. We would do well to examine this assumption here for it is closely related to the gift thesis. We just do basic actions, it is said: we cannot say how we do them; we cannot tell what we do first because there is nothing we do first. Notice that the claim is not simply that we do not know or have recipes for our basic actions; it is, rather, that there are none and can be none.

Given that a recipe is conceived of as a set of actions one must perform to perform the action in question there can be no doubt that basic actions have no recipes. It is important, however, to distinguish between there being a recipe for some action one performs and there being a recipe for acquiring the ability of which that action is an exercise. I have no recipe for raising my eyebrow; indeed there can be none. But there can be a recipe for acquiring the ability to raise an eyebrow; indeed I can give one. I say this distinction is important because it might be thought that an action for which there was no recipe would necessarily be a gift in that the absence of a recipe was incompatible with someone's acquiring the ability to perform the action. However an action for which there is no *performance recipe* might nonetheless have an *acquisition recipe*.

Now this distinction between two kinds of recipes might be protested on the grounds that for any given action the performance recipe and the acquisition recipe are identical and that if an action has no performance recipe it can have no acquisition recipe either. But surely this view is mistaken. The actions I performed in learning to swim may bear little resemblance to the ones I perform today in swimming. Thus there is no guarantee that the acquisition and performance recipes for a given action will be identical. Moreover, I do not have a performance recipe for raising my eyebrow although I do have an acquisition recipe.

To summarize, provided "recipe" is construed as "performance recipe," the recipe thesis of the standard view of basic actions is true: basic actions have no recipes of this sort. It does not follow, however, that basic actions are necessarily gifts since the lack of a performance recipe is compatible with the existence of an acquisition recipe.

Suppose now that it is maintained that basic actions not only have no performance recipes but have no acquisition recipes either. Is the gift thesis not made secure? My eyebrow raising case is in my judgment a clear counter-example to the claim that basic actions have no acquisition recipes. Moreover, I have checked with people who can wiggle their ears as a basic action and some of them claim to have learned how to do so and to have acquisition recipes for ear wiggling. But even if these examples are rejected, it should be clear that there are things we learn for which we have no acquisition recipes. Consider someone speaking his native language. No one doubts that he learned to do this yet we know relatively little about how he learned to do it and, so far as I know, we know no acquisition recipe for normal language learners.

---

[4] Melden's expression, *op. cit.*, p. 38.

[5] In "What We Can Do," p. 440, Danto cites wiggling one's ears as a basic action for some people. My calling raising an eyebrow an action ought not then to cause concern.

[6] Danto's term, "Basic Actions," in Brand (ed.), *The Nature of Human Action*, *op. cit.*, p. 263.

The lack of an acquisition recipe may simply reflect our lack of knowledge of the learning process. Thus it is possible that there could be an acquisition recipe for an action which at the present time has none. Indeed, it is possible that some actions whose acquisition recipes we do not know do in fact have such recipes. Thus, for example, it is possible that a person learning his native language does learn it by performing certain actions but that we have not yet discovered what these actions are: his action of speaking his native language would in such a case *have* an acquisition recipe although we would not *know* that recipe. But of course an action's having an acquisition recipe is not necessary for the action to be learned. An acquisition recipe, like a performance recipe, consists in performances of actions by the agent in question. It is, however, possible for a person to acquire the ability to do something without himself performing actions to acquire it: others may perform actions and may do things to him but their actions and the things they do to him are not actions of his. In cases of this sort the action would not have an acquisition recipe as these are conceived of here. Forms of conditioning, brainwashing, and the like would seem to involve just this: acquisition without acquisition recipes.

Recipes and examples aside, any thesis that requires a line to be drawn between gifts and what is learned or acquired would seem to be on shaky ground. On the one hand neither the notion of a gift, innate knowledge, or the like nor the notion of learning, acquiring some ability or the like is sufficiently clear to allow us to say with any confidence that a given action or ability is a gift or not. On the other hand, even if these notions were clarified sufficiently so that we knew what would count as a gift or as not a gift we would not—or at least we do not now—have enough information about the way in which our actions "emerge" to be able to decide whether they were gifts.[7] One ought not to put stock in a thesis like the gift thesis about basic actions until the results of a good deal more philosophical, psychological, and physiological research are in.

II. Some Puzzling Cases

If the gift thesis is untenable, then we can at least wonder if actions such as opening a door and knotting a tie which on the standard view are taken to be non-simple and to be clear cases of non-basic actions can be basic actions. In this section I shall take the first step in arguing that such things can be basic actions.

I play the first movement of a Schubert sonata on the piano. A proponent of the standard view of basic actions asks me "How did you play the piano?" I am tempted to say "I just did play the piano" because from my point of view I did just do it, but I know this will not satisfy him. So I say "I played the piano by moving my fingers" and I start to give a performance recipe: "I pressed my third finger right hand down on the *C*♯, my fourth finger right hand on the *D*, my right thumb on the low *C*♯. . . ." My action of playing the piano is clearly not basic for it includes or consists in perhaps thousands of other actions. Or does it? Is it really the case that in the context envisioned, namely my playing a part of a Schubert sonata with a fair though not extraordinary degree of accomplishment, my playing of each note is an action? To me this sounds implausible yet it is, presumably, the position to which the standard view of basic actions is committed.

G. Vesey has his readers consider the case of a little girl who acquires the ability to skip rope to music.[8] According to Vesey, if she were asked "How did you skip rope?" she would say "I just did." But of course this would not satisfy her interrogator. He would count each leg and arm movement as an action. Similarly if Karen Magnussen on the evening she won the Canadian Senior Women's National Championships had been asked, "How did you ice skate?" she would have been expected to say "By moving my legs." Each leg movement (her slip on the ice excepted, I suppose) would on this view qualify as an action.

---

[7] "Emerge" is used here as a neutral term—no suggestion of giftedness is intended.

[8] "Conditioning and Learning," in R. S. Peters (ed.), *The Concept of Education* (New York, 1967), pp. 61–72.

One wants to say "These aren't actions, they are just (mere) bodily movements" but given the framework accepted by proponents of the standard view of basic actions this will not do. A distinction between bodily movements and actions constitutes part of this framework but it is the distinction between one's arm moving and one's moving one's arm hence it cannot be invoked in these cases. Suppose your arm moves because of a spasm or because someone else pushes it. On the accepted view this would be a "mere" bodily movement not an action. But in playing the piano my fingers don't move as the result of some outside force or some inner push. I move my fingers. And in ice skating Karen Magnussen moves her legs. Given the accepted framework, there is nothing for these things to be but actions.[9]

Consider now a rather different sort of case. I'm writing at my desk. Let us grant that this is a non-basic action: I write by moving my arm and fingers, these are my basic actions. But while performing this action I wiggle my leg, I move forward in my chair, I scratch my head, I lean my elbow on the desk, I lean my head on my hand, etc., etc. In the course of performing an action one almost invariably does a lot of other things. Are these one's actions? Most of the things I do while sitting at my desk writing I'm not even aware of doing. If someone stopped me as I was shifting in the chair and told me what I was doing I might or might not agree with him. If he asked me what I was doing, I might or might not be able to tell him. In any case, my attention is not on these things and I do not realize that I am doing them unless something makes me aware of them. Again, these things do not fit the accepted notion of a mere bodily movement for I do them, they do not just happen to me. On the other hand, they seem a good deal more like things that do just happen to one, e.g., eye blinks, than like full-fledged actions, e.g., my writing at my desk.

My point in discussing these cases is to suggest that there are a lot of things we do in or while performing some action which do not themselves seem to be full-fledged actions yet are surely not "mere" bodily movements. Now I am relying here on my pre-analytic intuitions about what is and is not an action and am hoping my intuitions on this matter accord with my readers'. On the one hand I must rely on these for there is no adequate account or definition of action against which we can measure these cases. But on the other hand one wonders if I have the right to rely on these. Our intuitions fail us so often in determining if something is an action, do we have the right to have confidence in them now? (Is, for example, to sneeze an action? Or to breathe, to see, to checkmate, to think something, to say something?[10]) Moreover "action," as used by philosophers, seems to be more a term of technical philosophy than a term of everyday life. But then, what has one's intuitions to do with the matter? Supposing that an account of a term in ordinary language must accord with our intuitions about cases, accounts of technical terms surely do not have to meet this criterion.

What I think we are confronted with here is a matter of decision. Our pre-analytic intuitions, whatever their worth, are jarred if the sorts of things discussed here are classified as actions. We must then decide how we are going to treat these cases. We can if we want classify them together with actions and expand or loosen up our account of actions to accommodate them. Or we can introduce a category which falls somewhere between "mere" bodily movements and actions—a category in which we put those doings which are in some but by no means all respects like actions. I do not mean to suggest that these are our only alternatives. We can, for example, scrap the entire framework of actions, bodily movements, goings on, etc., and begin over, and there are surely other possibilities as well. The decision which confronts us arises in a context of theory development and there are therefore many options open to us.

III. Two Kinds of Doings

A number of philosophers have pointed out that not all of the things we may be said to do can be said to be our actions.[11] For example, a perfectly proper answer to the question "What is he doing now?" is "He's sleeping" and a perfectly proper answer to the question "What did you just do that got everyone so excited?" is "I won the race."

---

[9] Vesey in his discussion of the girl skipping rope is a bit misleading. He argues that the only answer one could give to "How does she do it?" would be in terms of nerve impulses, muscle contractions, and the like and points out that these are not things she does but are "goings on." *Ibid.*, p. 63.

[10] These examples are J. L. Austin's, "A Plea for Excuses," *Proceedings of the Aristotelian Society*, vol. 57 (1956–57), pp. 1–30. Reprinted in Alan R. White (ed.), *op. cit.*, pp. 19–42.

[11] For example Ruth Macklin, "Doing and Happening," *The Review of Metaphysics*, vol. 22 (1968), pp. 249–250; Alan R. White (ed.), *op. cit.*, introduction.

But if this is so—if, that is, there can be doings which are not actions—why can't such things as my playing *C*♯, etc., while playing a Schubert sonata and a champion skater's moving her legs be considered doings without thereby being considered actions? Of course actions, on the view of those who say that not all doings are actions, are themselves doings. But as long as there can be doings which are not actions we have a place to put those things we have just been discussing—things there is no doubt we do, yet things that seem so different from clear cases of actions.

There is of course one problem with this way of dealing with our "halfway house" cases. Moving one's leg while skating and shifting in one's chair while writing seem as far removed from things like sleeping and winning a race as they do from full-fledged actions. If we are to consider these cases of doings which are not actions, then, we must recognize that doings which are not actions are by no means all of a piece; indeed that there are two kinds of doings which are not actions—doings which are capable of being performed as actions and doings which are not. Presumably sleeping would fall in this latter category. Perhaps winning a race would too although I am less sure about its status vis-à-vis action. In any case moving one's leg, jumping, moving a finger, leaning forward—all doings which are not actions in the contexts envisioned above—would fall in the former category for they surely can be performed as actions in some contexts.

What is of course now needed is a clear basis for distinguishing between doings which are actions and doings which are not although they are capable of being performed as actions.[12] I shall propose here a basis, or at least a partial basis, for the distinction and shall consider what might be thought to be a crucial objection to it.

The "halfway house" cases I have been discussing are cases in which the doer's attention is on something else, his effort is directed elsewhere. The little girl jumps and turns the rope without attending to these elements of her performance. Her attention is focused on the overall task of skipping rope and not on some part or parts of it. Even if, when writing, one's attention is focused on moving one's hand in the proper way or on holding the pencil, it is not (at least not in the case discussed above) focused on moving forward in one's chair or scratching one's head. Were I to skip rope, my attention would be focused on both the overall task and some parts of it, e.g., jumping. When some people, notably but not only beginners, ice skate their attention is focused on every step or glide; when I skate I do not attend to each step but I do without a doubt attend to certain elements of skating, e.g., stopping.

What I propose is that a distinction be drawn between those doings to which a person attends and those doings to which he does not attend. The doings to which he does not attend are "mere" doings, i.e., they are not actions. This is not to say that the doings to which he does attend are necessarily actions. Attending to what one is doing is a necessary condition for that doing's being an action.[13] It may also be a sufficient condition but this is not an issue that needs to be decided here. The halfway house cases I have presented are "mere" doings, I submit, because the doer is not attending to what he is doing. Supposing he were attending, we might feel certain that these were his actions. But if we still did not—if, that is, some further ingredient of action were lacking, my case against the standard view of basic actions would not be affected. The standard view would still count as actions things that ought not to be so counted. Moreover, if something over and above attending were needed to make a doing qualify as an action, my thesis that actions which the standard view takes to be non-simple can be basic actions would in no way be undermined.

Now the obvious objection to this proposal is that some things are done unwittingly or unintentionally or absentmindedly (or accidentally or mistakenly or involuntarily, etc.) and that the criterion of attending rules them out. We must consider, therefore, whether our account does rule such things to be non-actions. If it does, we must decide whether this is a fatal defect. Our task is made difficult because there are so many varieties of doings which might be affected by our proposal and because a comprehensive, systematic account of them is not available. Since it is evident that an attempt to provide such an account is well beyond the scope of this paper, what I shall do is examine a few representative cases of things done unwittingly, unintentionally, etc., and see what can be concluded from them.

---

[12] As well as, of course, a clear basis for distinguishing doings from other things and a clear basis for distinguishing between doings which although not performed as actions are capable in a particular context of being actions and doings which are not capable of this "shift" from "mere" doing to action.

[13] For discussions of this ingredient of action see Melden, *op. cit.*, p. 134ff. and Brian O'Shaughnessy, "Observation and the Will," *The Journal of Philosophy*, vol. 60 (1963), pp. 367–392.

Consider someone signing a document he does not want to sign because there is a gun in his back. The account I am proposing here accords with our pre-analytic intuitions: it does not disqualify this doing as an action for in the context envisioned the person attends to what he is doing although he is doing it against his will. This sort of doing is often called involuntary. It is quite different, however, from another sort of involuntary doing, e.g., someone sighing without realizing it or meaning to. This latter would not on the present account of action qualify as an action but surely this too accords with our pre-analytic intuitions.

Now consider someone's reaching across the table to get the ketchup and in so doing knocking over a glass of milk. Let us suppose that there is no question that his reaching across the table is an action. What about his knocking over the glass? Presumably his attention was not directed to this; indeed had it been he would not have knocked the glass over. His knocking over the glass on my account is not then an action; it and countless other accidental or unintentional doings are ruled out by my attending condition. Is our account to be condemned? I think not, for it is open to question if accidental or unintentional doings of this sort are actions.

Some philosophers seem to think they are actions. Joel Feinberg, for example, thinks we can "puff out" an action to include an effect: supposing Jones in opening the door caused Smith to be startled and Smith suffered a heart attack and died; we can puff out Jones's action of opening the door to Jones's startling Smith and even Jones's killing Smith.[14] Other philosophers, however, deny that our diner's knocking over the glass or Jones's startling Smith, let alone killing Smith, are actions. Accidental or unintentional doings of this sort are to be understood, they argue, as actions (our diner reaching for the ketchup, Jones opening the door) and their unintended effects or consequences.[15] We have here cases about which even the experts disagree. They are then *at best* unclear or borderline cases of action. But then the objection to attending as a necessary condition of action on the grounds that it rules out these cases ceases to be an objection. Unclear cases will fall on one side of the line or the other given some attempt to clarify the notion of action, but no one side is necessarily preferred.

There are other sorts of doings which my account rules out as actions. Consider my pouring the orange juice into my coffee cup at breakfast although I intended to pour the juice into my glass. Clearly I was not attending to pouring the juice into my coffee cup. This and other absentminded doings like it are ruled out as actions on the account I am giving here.[16] So too habitual doings such as my putting on my right shoe before my left shoe are ruled out as actions by my account. But neither pouring my orange juice into my coffee cup nor putting on my right shoe before my left shoe seems to be a clear case of an action. Philosophers of action have not to my knowledge examined cases like these to the extent that they have examined cases of accidental doings like the ones discussed above. Yet they would be bound to disagree about them if only because they disagree over whether actions necessarily involve intention.

Consider finally someone's delivering a package which unbeknown to him contains a bomb. Assume that he attends to delivering the package; indeed that delivering the package is an action of his. Is delivering the bomb an action of his? Again we seem to be confronted with an unclear or borderline case of action. On the one hand our man doesn't intend or want to deliver the bomb nor does he realize he is delivering a bomb. In important respects this is like the case of the diner who knocked over the glass. Some philosophers might even want us to understand his doing (delivering the bomb) as an action (delivering the package) and an unintended event (a bomb being delivered)—although of course the event would not as in the case of the spilled milk be a causal consequence of the action. On the other hand his delivering the package *is* delivering the bomb. Can it seriously be held that the latter is not an action although the former is?

Unlike the other borderline cases we have discussed, it is not clear that the bomb case falls outside the realm of action on the account being offered here. In one sense our man is clearly not attending to what he is doing, i.e., delivering the bomb—he does not realize or believe or think he is delivering a bomb. But in another sense he is attending to this doing—he carries the package containing the bomb carefully, makes sure he puts it in the hands of the person he is supposed to, etc. Until the notion of attending to what one is doing is itself clarified delivering the bomb must be considered an unclear or borderline case of action relative to the present proposal. But this does not constitute a reason for abandoning our attending condition; it simply points up the need for further clarification. Given an adequate account of attending, delivering the bomb and doings like it may turn out to meet our attending condition in which case the objection that our proposal rules out too much will not be relevant; or these doings may turn out not to fulfill the attending condition in which case the objection, although relevant, will come to naught since they were initially borderline cases and, as we have seen, there is no one direction in which borderline cases must fall.

---

[14] "Action and Responsibility," in M. Black (ed.), *Philosophy in America* (Ithaca, 1965), p. 146. Reprinted in Alan R. White (ed.), *op. cit.*, pp. 95–119.

[15] See, for example, Jack Meiland, "Are There Unintentional Actions?", *Philosophical Review*, vol. 72 (1963), pp. 377–381.

[16] Not all absentminded doings are ruled out by my account however, e.g., someone's taking the subway home forgetting that he had driven his car to work and had intended to drive it home.

In sum, the cases we have examined here provide us with good reason for rejecting the claim that in ruling out unwitting, unintentional, etc., performances as actions our account rules out too much. Our account seems to accord with pre-analytic intuitions about clear cases. And insofar as it rules out unclear or borderline cases it is within its rights.

Interestingly enough, in his essay "On the Characterization of Actions," Nicholas Rescher draws a distinction very much like the one I have been drawing but for a very different purpose.[17] I take note of it here because it seems to me to provide independent support for the present proposal. Rescher's concern is with the question of the infinite divisibility of action. Cannot every action be divided into further actions? For example, if I move my right foot forward one step, don't I first move it forward one-half step and so on *ad infinitum*? But if this is so, can the distinction between basic and non-basic actions be upheld? Can there be componentless actions? Rescher blocks this criticism by arguing that moving one's foot forward half a step is in context not envisioned as an action but is rather a potential action: it is something one could do if the doing of it were "on one's mind" and one wanted to do it.[18]

Rescher cites also the action of lifting a spoonful of sugar out of a sugar bowl. He points out that in performing this action one does lift out the component grains. But action has its intentional aspect, he says, and if the spoonful and not the grains are on the agent's mind, then the action of lifting out the spoonful is not composed of the action of lifting out the grains.[19] Although we may not be able to attribute to Rescher the precise distinction I have introduced here, it is worth noting that his reason for calling moving one's foot one-half step forward and lifting the grains out of the sugar bowl potential actions is that they were not *on one's mind.* This is not too different, if indeed it is different at all, from my claim here that some things which can in certain contexts be actions are not our actions because we are not attending to them.

It should be noted that Rescher's intent in denying that moving one's foot forward one-half step and lifting grains out of the sugar bowl are actions in the context envisioned was to save the basic/non-basic actions distinction. Yet advocates of the standard view of basic actions count as actions all sorts of things one does which one does not attend to and which I should think Rescher would have to admit are not on one's mind—e.g., moving one's leg in skipping rope. Indeed, they must count these things as actions if they are to maintain, as they do, that the candidates for basic action are few, that moving a limb is a paradigm case of a basic action, that in a word only those actions which they take to be simple actions can be basic actions. For if the things in our performances which we do not attend to or are not on our minds are not counted as actions, the basic action door is open to all sorts of actions which they take to be complex—ice skating, playing the piano, opening a door by turning a knob, and so on. I myself see nothing wrong with this, but one can well imagine some proponents of the standard view of basic actions protesting that if this is a consequence of Rescher's attempt to save the basic/non-basic action distinction, then he had better not have bothered for in trying to save the distinction he has in effect destroyed it.

IV. A Revised Account of Basic Actions

Whatever Rescher's position may be, it seems to me that these things can be performed as basic actions. To be sure, if we conceive of actions such as these as being capable of being performed as basic actions, the distinction between basic and non-basic actions will in some respects be a changed one. But it will not be destroyed; nor will its significance in the theory of action be undermined.

---

[17] In Myles Brand (ed.), *The Nature of Human Action, op. cit.*, pp. 247–254.

[18] *Ibid.*, p. 252.

[19] *Ibid.*, p. 253.

Let us, then, adopt the distinctions we discussed above between doings which are actions and doings which are not and in the case of the latter between those doings which are capable of being actions and those which are not. Given that a basic action is one which is not done by doing some other *action*, a person skipping rope can be said to be performing a basic action provided that the things he is doing in skipping rope, e.g., jumping, etc., are not themselves actions. And similarly with ice skating, playing the piano, opening a door by turning the knob, and so on. Now I take it that in the case of Vesey's little girl, the things she does while skipping rope are doings which are not actions for she skips rope smoothly, pays attention to the global task and not the subsidiary ones of, e.g., getting over the rope at each turn without tripping. Compare her performance with mine. Skipping rope to music would not be a basic action for me. I would have to focus on turning the rope and on jumping. Other things being equal these would be actions of mine which were included in my action of skipping rope.

Note that on the view I am proposing, what is a basic action in one context may be a non-basic action in another context. This relativity to context is of course a feature of the standard account too. But furthermore, and here my account differs from the standard account, if an action is performed as a non-basic action, the reasons for its being a non-basic action—that is to say, the actions it includes or consists in—will differ from one context to another. Consider the difference between a beginner playing a rather simple piano piece, my playing that piece, and an expert's playing it. For all of us playing the piece is a non-basic action. But for the beginner, the actions included in the overall action of playing the piano are the actions of moving his fingers; for me the component actions are the actions of playing the left hand runs and the trills; for the expert the component actions are the actions of pedaling and phrasing. Thus one cannot tell *a priori* what the component actions of a non-basic action are: one man's actions will be another man's "mere" doings and still another man's "mere" happenings or goings-on. To tell what actions are components of a non-basic action one must again determine what the agent is attending to.

I should perhaps forestall one possible misunderstanding by stressing that I am not saying that a particular action is a basic action from the performer's point of view and a non-basic action from some other point of view or that some component of a non-basic action is an action from the performer's point of view and a "mere" doing from some other point of view. On the contrary I have tied down what anyone, performer or otherwise, is to take as a basic action or as an action component of a non-basic action to the performer's point of view in the sense that the judgment as to basic or non-basic, action or "mere" doing, rests at least in part on what the performer is focusing his attention on. Thus, granted that the little girl skipping rope is focusing solely on the "global" performance, her action of skipping rope is from her point of view and ours a basic action; and granted that the expert pianist attends to his pedaling and phrasing and that these are his actions, they are his actions from his point of view and ours.

The standard view of basic actions rules out *a priori* as basic actions such apparently non-simple actions as skipping rope, playing the piano, and opening a door by turning a knob. On the present view such actions can be basic actions but it is an empirical question whether in fact such actions are ever performed as basic actions. I call skipping rope, playing the piano, and opening a door *apparently* non-simple because the notions of simple and complex are relative. A thing is not simple or complex *per se*; rather it is considered to be simple or complex relative to some interest or purpose. The standard view of basic actions ignores this relativity: basic actions are in all contexts and from all points of view assumed to be simple and non-basic actions are in all contexts and from all points of view assumed to be complex. Arthur Danto, for example, says:

> A basic action is perfectly simple in the same sense in which the old "simple ideas" were said to be: they were not compounded out of anything more elementary than themselves, but were instead the ultimately simple elements out of which other ideas were compounded.[20]

To be sure, basic actions have no action components. By definition they cannot. If this is all that is meant when basic actions are called simple actions, the simplicity thesis of the standard view is unobjectionable. Given this sense of "simple action" basic actions on the view I am proposing are simple actions. Given this sense of "simple action," however, all sorts of actions which according to the standard view are non-simple actions can on my view of basic actions be simple for all sorts of actions which are assumed by the standard view to be non-simple can be performed as basic actions, i.e., as actions which have no action components. Insofar, then, as simple actions are merely actions which have no action components, the view of basic actions suggested here requires no conceptual divorce between basic actions and simple actions. It does, however, admit as candidates for basic actions, actions which the standard view takes to be non-simple.

---

[20] "Basic Actions," in Brand (ed.), *The Nature of Human Action*, *op. cit.*, pp. 264–265.

The standard view, however, holds that basic actions are simple in the sense of being indivisible. Now it does not follow from the fact that basic actions have no *action* components that they have *no* parts or components. On the view of basic actions I am proposing here, a basic action can have doings as components. But even on the standard view there is nothing preventing a basic action from having as its components such things as goings on or happenings, rules or social customs, purposes or reasons. To say this is not to say that a basic action is *reducible* to such components: my house has rooms as its parts, but this is not to say that my house is reducible to rooms or composed solely of rooms.

The distinction between basic and non-basic actions is intended to be neutral with regard to the definition of action and in view of the fact that the definition of action is a very controversial issue in philosophy this is as it should be. However, the view that basic actions can have no components seems to interfere with such neutrality: in effect it rules out in advance definitions according to which actions have non-action components. From a theoretical standpoint then it certainly seems desirable to separate the basic/non-basic and the simple-complex distinctions, except of course in the narrow sense of "simple" according to which a simple action merely has no action components.

In arguing that basic actions are not necessarily indivisible I do not mean to be suggesting that they are necessarily complex. As I have said, a thing is not simple or complex *per se* but is considered to be simple or complex relative to some interest or purpose. One person may view an action which is performed as a basic action as complex; another may view it as simple. So too, one person may view an action which is performed as a non-basic action as complex while another views it as simple. Suppose one's interest is in the pleasure or pain a given action will produce. The action may well appear simple, i.e., indivisible, a unity, whether it be performed as a basic or a non-basic action. If, however, one's interest is in physiology, a given action, be it performed as non-basic or basic, may well appear complex.

We gain neutrality with respect to the question of definition and at the same time recognize a dimension of actions which we have tended to ignore if we effect a conceptual divorce between basic actions and simple actions. Armed then with two distinctions—basic/non-basic actions and simple-complex actions—rather than one, let us look again at cases.

Presumably the little girl skipping rope conceives of her performance as a simple action: her skipping rope is, therefore, from her standpoint a simple action and of course it is also a basic action. A spectator, however, might conceive of her action as complex: "Look at all the things she is doing in skipping rope!" he might exclaim. Her skipping rope is from his standpoint a complex action albeit a basic action. The same spectator might also view my skipping rope, by hypothesis a non-basic action, as complex and I would surely agree with him about its complexity. Thus spectator and performer may agree in a given case on the simple-complex question although clearly they don't have to. Nor must two spectators take the same stand on the complexity or simplicity of a given action. If while I am skipping rope the rope hits a bystander, he and his relatives may take my skipping rope to be a simple action, a unity, which had an unfortunate consequence and not a complex whole composed of a number of elements. This could be true for the injured bystander, I should add, whether my skipping rope was a basic action or a non-basic action.

V. Conclusion

I can well imagine someone protesting at this point that my revised account of basic actions so changes the original distinction between basic and non-basic actions that I have no right even to call my basic actions *basic*. On the standard view, basic actions are basic in that they are the "starting point" for the performance of other actions; they are, so to speak, the building blocks out of which other actions are constructed. Our open door policy toward basic actions does not change things in this regard. Basic actions can serve as a "starting point" for performances of other actions on my account as well as on the standard view. Consider, for example, a person's action of ice skating. On my view it is possible for him to ice skate as a basic action although on the standard view it is not. Suppose now he does—he ice skates without performing any other action. Can his ice skating serve as a building block for some other action? I see no reason why it cannot; for example, ice skating can serve as a starting point or building block for a person's action of playing hockey. The "size" of a basic action can be much greater on my view than on the standard view, but "size" is no bar to an action's being *basic* in the sense at issue.

Can any action, then, be performed as a basic action? I have discussed skipping rope, playing the piano, and ice skating at length, and it may be wondered if the actions that advocates of the standard view of basic actions most often discuss—e.g., moving a stone by kicking it, opening a door by turning the knob, signalling by raising one's arm—can be performed as basic actions. These are the things that are normally held to be beyond the pale, i.e., to be intrinsically non-basic. Compare a child's opening a door by turning a knob and an adult's opening the same door. The child attends to turning his hand, to turning the knob, to pulling the door. For him this is clearly a non-basic action. An adult (unless, e.g., he is showing the child how to open the door) attends to none of these things—he just opens the door. For him this is clearly a basic action. So too with moving the stone by kicking it, signalling by raising one's arm, and, so far as I can see, any other action. There may be limits on what in fact can be performed as a basic action but such limits can not be set *a priori*. Some actions, for example, may be too complex to be performed without our attending to some components, but if this is so this would be a function of our physical and psychological limitations rather than a function of the inherent nature of actions.

Radcliffe Institute

Received September 22, 1970

American Philosophical Quarterly
Volume 9, Number 1, January 1972

VII. MEANING AND ILLOCUTIONS

DAVID E. COOPER

I

A question that has been much discussed in recent years is "Can the meanings of sentences and words be explained by reference to illocutionary acts?" Even where the question has not been raised in this explicit form, a good deal of literature seems to provide answers to the question. The student of the literature, moreover, cannot help being impressed or depressed by the bewildering number of apparently conflicting answers given. It seems that at least four, apparently inconsistent, answers have been given to the question. I shall briefly state these, and refer to them as theses (1) to (4). They are:

(1) There are illocutionary acts, and meanings can be explained by reference to them. (W. Alston)
(2) There are illocutionary acts, but any attempt to explain meanings by reference to them would be circular. (D. Holdcroft)
(3) There are illocutionary acts, but these are quite distinct from meanings, and so presumably cannot be used in explaining meanings. (J. L. Austin)
(4) There are no such things as illocutionary acts, and so meanings cannot be explained by reference to them. (L. J. Cohen)

This paper is an essay in reconciliation. I want to argue that none of the four theses, despite appearances, are inconsistent with any of the others. Once each thesis is explicated, and ambiguities are removed, it can be seen that there is a thesis which the proponents of the above four could all agree upon. I shall not, apart from a few very brief comments at the end, try to assess that thesis. My aim is simply to show that the four are consistent, hoping thereby to lend a little order to the discussion of meaning and illocutions. In Sect. II I try to show that theses (1) and (2) are consistent, since Alston and Holdcroft are in fact dealing with separate questions. In Sect. III I try to show that theses (3) and (4) are consistent since, in a good sense, the difference between them is a purely verbal one. Finally, in Sect. IV, I try to show that theses (1) and (3) are consistent, once a certain type of distinction between the general and the particular is made. If each of the sections is successful, I shall have shown that all four theses are mutually consistent.

II

Thesis (1) is argued for by William Alston in a number of places. According to him, to say that one linguistic expression *E*1 means another linguistic expression *E*2 is to say that *E*1 and *E*2 have the same use. He then claims that the relevant sense of "use" which is involved here is that of "illocutionary force." On this basis he erects definitions of meaning in terms of illocutions. Where *S*1 and *S*2 are sentences, we can say:

> *S*1 means *S*2 =df. *S*1 and *S*2 have the same illocutionary act potential.

Where *W*1 and *W*2 are sub-sentential linguistic expressions, we can say:

> *W*1 means *W*2 =df. *W*1 and *W*2 can be substituted for each other in a wide range of sentences without altering the illocutionary act potentials of those sentences.[1]

That is, two sentences mean the same if both can perform the same range of illocutionary acts; and two sub-sentential expressions mean the same if one can be substituted for the other without altering the ability of the sentences in question to perform certain illocutionary acts. Alston, in fact, refines these definitions to account for multi-vocality, but such refinements need not concern us. Let us assume we are dealing with a completely disambiguated language.

Thesis (2) is argued for by David Holdcroft in his criticism of Alston's views. According to him the enterprise of defining meanings in terms of illocutions must meet a certain criterion of adequacy. Let *K* be a class of expressions in a language *L* whose meanings, supposedly, can be explained solely by reference to illocutionary acts. The criterion of adequacy is

---

[1] William Alston, *Philosophy of Language* (Englewood Cliffs. N.J., 1964), pp. 36–37.

> that in the case of each speech act referred to in explanation of the meaning of a member of *K*, there is a way of telling what speech act has been performed which does not involve making use of semantic information about *L*.[2]

The point is: unless one can know what illocutionary acts are performed by a sentence *S* independently of already knowing the meaning of *S*, then any attempt to explain that meaning in terms of those illocutionary acts will be circular. Or, as Holdcroft puts it,

> the claim that T.I.A. (the Theory of Speech Acts) can be used to explain the meaning of all (or most) words is uninteresting unless some restriction is placed on the use of semantic information in conjunction with T.I.A.[3]

Holdcroft argues that the above criterion of adequacy cannot be met. While one can know what speech acts of a relatively unspecific sort are being performed independently of already knowing the meanings of the sentences employed, this is not so once the speech acts are specified in detail. We might know that a person is giving a command without knowing what his utterance means, but once we give a full description of what he is doing—such as commanding Johnny to close the door quickly—then "it is obviously less plausible to hold that there is an independent way of telling that an act, so described, has been performed."[4] Clearly, moreover, the descriptions of illocutionary acts which would be relevant to explaining meanings must be of the full, specific type. Otherwise we should render any number of non-synonymous sentences synonymous on the grounds that, for example, they are all used to command. But illocutionary acts, once fully described, can no longer be identified independently of already understanding the sentences used to perform them.

Holdcroft thinks that it is possible independently to identify illocutionary acts only where two conditions hold. The conditions are (1) that the sentences are, in Ziff's sense, non-determinate: that is, the sentences must not be context-free. (E.g., "John hit(s) James at 2:30 P.M. on November 1 1970" is determinate relative to "He hit him then.") And (2) that the illocutionary acts performed are not essentially verbal. That is, it must be the case that they could have been performed by means which are not verbal, and not parasitic upon verbal means (in the way that a nod of the head is parasitic upon a statement of agreement). For if (1) context cannot inform us as to what illocutionary act the speaker is performing, and (2) there is no non-verbal means of performing the act with which we might associate the utterance, then we can only recognize what illocutionary act is being performed if we already know what the sentence means. In such a case appeal to illocutionary acts to explain meaning would be circular. Now according to Holdcroft, the number of utterances meeting the above conditions are

> so small that it would be quite implausible to suppose that any set of operations could be specified on the members of *K* which would yield the meaning of all or most words of the language of which *K* was a subset of sentences.[5]

It is not my intention to criticize Holdcroft's arguments but, rather, to cast doubt upon their relevance to Alston's position. There are at least two prima facie reasons for doubting their relevance. In the first place, and less importantly, Alston claims to be following Austin's account of illocutionary acts—so he can hardly be unaware that most of the acts in Austin's list are essentially verbal. Austin, indeed, makes that clear. Consequently, Holdcroft's second point against Alston is one which it is hard to think Alston would want to deny. Secondly, and more importantly, Alston is quite happy to admit that independent identification of a speech act from the sentence used to perform it, in Holdcroft's sense, is impossible. He admits that one cannot give the meaning of a sentence. In terms of illocutionary potential, "in the theoretical sense of being able to say what its potential is."[6] Again, he makes it clear that "we do not presently have in our conceptual repertoire adequate resources for saying what the use of a word is."[7] I take these and other passages as admissions that the relationship between a sentence *S* and various illocutionary acts it may be used to perform is not such that one can independently give full descriptions of those acts apart from semantic knowledge about the sentence. Alston, then, admits what Holdcroft regards as the major defect in his enterprise.

---

[2] D. Holdcroft, "Meaning and Illocutionary Acts," *Ratio*, vol. 6 (1964), pp. 128–143. Also in *Theories of Meaning*, ed. by Parkinson (Oxford, 1968), p. 170.

[3] *Ibid.*

[4] *Op. cit.*, p. 174.

[5] *Ibid.*, p. 176.

[6] *Op. cit.*, p. 39.

Unless we are to suppose that Alston is quite happy to contradict himself, it would be wise to look again. Perhaps Holdcroft has misunderstood his position. It seems to me that he indeed has. Holdcroft treats Alston as if the latter were saying: to specify the meaning of a sentence is to describe the set of speech acts it is capable of performing. But what Alston says, however, is this: to specify the meaning of a sentence *S*1 is to cite another sentence *S*2 which has the same illocutionary act potential as *S*1. He says, for example, that

> saying what the meaning of an expression is, is equivalent not to saying *what* its use is, but rather to saying that it has the *same use as another expression.*[8]

If Alston had been trying to define meaning in terms of what illocutionary acts a sentence performs in the way Holdcroft supposes, then perhaps he would be open to the objections mentioned. Since this is not what he does, I do not see how the objections carry weight against him. Alston, in fact, does admit a sense of meaning, or "knowing the meaning of," which cannot be non-circularly described by reference to illocutionary force. This is the

> practical, know-how sense of being prepared to use it (a sentence) to perform certain illocutionary acts and not others, and of being able to recognize misuses.[9]

But Alston's prime concern is to explain what it is to specify or give the meanings of sentences; and this is to give other sentences having the same illocutionary force. Of course, such specifications will be useless unless the addressee "knows the meaning of" the sentences used in explanation in the "practical, know-how sense." Still, I cannot see how Holdcroft's arguments operate against this view of meaning-specification.

Somebody—Holdcroft perhaps—might want to object that the charge of circularity still holds against Alston. For how can I informatively explain the meaning of *S*1 unless I can independently specify the set of illocutionary acts which it, and *S*2, are both capable of performing. If I cannot first know what the illocutionary acts are, how can I know that *S*1 and *S*2 are identically related to them, and synonymous in virtue of that relation? This objection does not succeed. It is, of course, true that very often we show two items to be identical by discovering some third item to which they are identically related. Thus I might show that the President and my next door neighbor are the same person because both possess the only platinum-plated Cadillac in America. However, informative identities can be asserted where it is not the case that the third term, in virtue of which two items are identical, is separately identifiable from the two items. I might say, in reply to the question, "Who is the most powerful man in the country?", that this person is the President, in virtue of the powers and functions of the President. Now here it need not be the case that the powers and functions referred to can be described fully without mentioning that they are Presidential powers and functions. That is, it might be essential to them being the powers and functions they are that they are the President's. This fact, as far as I can see, does not prevent my telling someone that the most powerful man is the President from being informative, or from explaining what makes him the most powerful man. Nor, as far as I can see, does the fact that *S*2 cannot be separately identified from certain illocutionary acts (in Holdcroft's sense) prevent the claim that *S*1 means *S*2 from being informative. For, after all, if I can use *S*2 successfully then, whether or not I can describe the acts performed by it in Holdcroft's independent manner, knowing that I can use *S*1 in the same way is surely significant information to me. (I do not want to claim that there is any further analogy between the President example and the case under discussion than the one I have drawn).

In fact, it is not transparent what Holdcroft means anyway when he insists that illocutionary acts, under full descriptions, cannot be separately identified from the sentences used to perform them. No doubt if one could say nothing at all about the acts except that they were the acts performed by a sentence *S*2, then it would be no explanation of *S*1's meaning to say that it meant the same as *S*2.

---

[7] W. P. Alston, "Meaning and Use," *Philosophical Quarterly*, vol. 13 (1963), pp. 107–124. Also in *Theories of Meaning*, ed. by G. H. R. Parkinson (Oxford, 1968), p. 159.

[8] "Meaning and Use," *op. cit.*, p. 159 [my italics].

[9] *Philosophy of Language*, *op. cit.*, p. 39.

But we must be careful about this question of what we can say about the illocutionary acts. What may well be true is that there is no way of describing the open class of speech acts performed by *S* beyond saying that they are the speech acts performed by *S*—just as, perhaps, one can say nothing about all possible games except that they are games. However, it does not follow that we cannot say more about particular illocutionary acts performed by *S*. For I might say of a particular illocutionary act that it is the act which, given conditions *X*,*Y*, . . . *N*, is performed by uttering the sentence *S*, or *S′*, or *S″*. So to tell me that *S1* means *S2* is not to tell me *simply* that it can be used whenever *S2* may be used. It might also tell me something about the conditions under which it may be used, and about certain other expressions it may be used instead of. Somewhat analogously, I no doubt convey little information when I say that the most powerful man is the President unless I can say something about the President's powers other than that they are the President's powers. But usually, of course, I can say more than this about certain of those powers—even if any complete description of them must make reference to the fact that they are presidential powers.

I conclude, then, that thesis (2) in no way conflicts with thesis (1), once the latter is understood in the manner Alston means it to be understood. Treated as an account of what it is to specify the meanings of sentences in terms of other sentences by virtue of identity of illocutionary potential, it is left unaffected by Holdcroft's arguments.

III

(It would be wise, for the moment, if the reader would forget the discussion in Sect. II. The relation between that section and this will be explained in Sect. IV.)

Thesis (3) is that the illocutionary act performed by a sentence is quite distinct from its meaning and, so presumably, cannot be used in explanation of the meaning. This thesis can be ascribed to J. L. Austin. According to him, meaning is an aspect of the locutionary act, namely the "rhetic" aspect. More exactly, to perform a rhetic act is

> generally to perform the act of using that rheme or its constituents with a more or less definite "sense" and more or less definite "reference" (which together are equivalent to "meaning").[10]

Now locutionary acts are distinct from illocutionary acts, and moreover pretty obviously distinct in a way in which the difference between illocutionary and perlocutionary acts is not always so obvious. Thus one can usually know what locutionary act a person performed without having the slightest idea what illocutionary act he performed. Austin, unfortunately, does not expand upon what he means by "sense" and "reference," which constitute "meaning," beyond saying that he is using them "on the strength of current views."[11]

L. J. Cohen, in a criticism of Austin's views, claims that "the concept of illocutionary force . . . is empty."[12] Consequently thesis (4)—that there are no illocutionary acts to appeal to in explanation of meaning—can be ascribed to Cohen. According to Cohen, no distinction can be made between meaning and illocutionary force. As we shall see, however, this is not to agree with Alston that the concept of illocutionary force can explicate meaning. Cohen, though, does not dispute many, if any, of the major points J. L. Austin makes in connection with illocutions. He agrees, for example, with Austin's demotion of statements; and with his doctrine of infelicities; and with the theory of speech acts when "viewed in the light of the support it offers to the campaign against descriptivist, or fact-value oversimplification."[13] In fact the reader of Cohen's article begins to wonder quite what is at issue between the two philosophers. Cohen, conveniently, places such puzzlement into the mouth of an imaginary objector. The "objector" says:

> But you yourself state the core of Austin's theory when you say that the illocutionary force of an utterance is that aspect of its meaning which is either conveyed by its explicitly performative prefix, if it has one, or might have been so conveyed by the use of such an expression. The only difference is that Austin made his point clearer by using the word "meaning" in a narrower sense, so that force stood out as something co-ordinate with meaning rather than as one special form or aspect of it.[14]

---

[10] J. L. Austin, *How to Do Things with Words* (Oxford, 1962), pp. 92–93.

[11] *Ibid.*, p. 148.

[12] L. J. Cohen, "Do Illocutionary Forces Exist?," *Philosophical Quarterly*, vol. 14 (1964), pp. 118–137. And in *Symposium on J. L. Austin*, ed. by K. T. Fann (London, 1969), p. 420.

[13] *Op. cit.*, p. 441.

[14] *Ibid.*, pp. 439–440.

Certainly it does appear that the sole difference between Austin and Cohen is that there are certain clarifications of utterances which Austin would describe as clarifications of illocutionary force, but which Cohen would describe as clarification of an aspect of meaning. They agree, on all other counts, as to the nature of the clarifications. For example, Austin would say that clarification as to whether an utterance is intended as praise or blame would be illocutionary clarification. Cohen, on the other hand, says:

> When you remark "In saying that I was praising his memory, not his intelligence," you are obviously clarifying meaning. So why not also when you remark "In saying that I was praising, not blaming him"?[15]

It does appear, too, as if the question we began with, "Can the meanings of sentences and words be explained by reference to illocutionary acts?", can be asked either in Austinian language, or in Cohen's language—and this would make no difference to how we should set about answering it. The question might be posed in either of the forms:

1. Can the illocutionary aspect of utterances explain the "rhetic" aspect? (Austin's)
   or
2. Can that aspect of meaning which is or might be conveyed by the use of explicit performative prefixes explain that aspect of meaning which cannot be so conveyed? (Cohen's)

And, it seems, that if Austin must answer his question negatively (as I am assuming he must), then such an answer would have to be given by Cohen in his terms. Indeed, it may seem obvious that Cohen is committed to a negative answer. For how can that aspect of meaning which can be conveyed by explicit performative prefixes explain that aspect which cannot? More of that in the next section.

I could leave things there. For I think it is clear that, for the sake of the discussion in this paper at any rate, it does not matter whether we employ Austin's or Cohen's manner of speaking. Since, as far as I can see, both agree upon the substantial points Austin makes in connection with speech acts, and both would answer our question in the same way. However, it would be unfair to Cohen to say that the only difference between what he says and what Austin says is a verbal one—for Cohen insists that it makes a "vital difference" which way of speaking we adopt.[16] He has a number of objections against Austin's classification. The main ones, I think, are these:

1. Austin's insistence that meaning and force are distinct leads him into trouble over explicitly performative utterances, such as "I protest that. . . ." For it seems he is committed to saying that the meaning of such an utterance is confined to the expressions which would fill the blanks. If the meaning was not so confined, it is difficult to see how meaning and force could be distinguished. However, such a conclusion is unacceptable. For surely "I protest," when used as a self-sufficient sentence by itself has a meaning; and surely it is implausible to suppose it loses that meaning when placed at the beginning of a longer utterance. If so, the meaning of such an utterance cannot be confined to what fills the blanks, and so meaning and force cannot be distinct.

2. "Austin's co-ordination of force with meaning wrongly suggests that the dissimilarity between reference and sense is somehow less than that between force, or performative function, and sense. . . ."[17] That is, if it is permissible to lump sense and reference together under the heading of "meaning," then surely it is at least as permissible to include force in the bundle as well and regard it as part of meaning.

3. Cohen points out that many clarifications of utterances which Austin would describe as clarifications of force can quite properly be described as clarifications of meaning. For example, to explain that one was praising and not blaming is to explain meaning.

Since, in the next section, I am going to relate thesis (1) with thesis (3); since, that is, I am going to prefer Austin's way of talking, I should say, very briefly, why I do prefer that way, and why I do not accept Cohen's objections. I take them in order:

1. This is a complicated point. I shall mention only two considerations. First, I do not see quite what Cohen means when he says that in the case of explicit performative utterances at any rate, meaning and force can definitely not be distinct. Surely, for example, I can wonder whether "I protest . . ." is being said as a joke, as an insult, or as a protest. If so, why should I not know its meaning, but still be in the dark as to what illocutionary act is being performed. At any rate, I do not see that the

---

[15] *Ibid.*, p. 428.

[16] *Op. cit.*, p. 440.

[17] *Op. cit.*, p. 440.

supposed non-distinction between meaning and force is any more obvious in the case of explicitly performative utterances than in others. Secondly, I refer the reader to Mats Furberg's reply to Cohen in which he argues, successfully I think, that the role of "I protest!" when used as a self-sufficient utterance is different from its use as a performative prefix. If that is so then Cohen's claim that "I protest . . .," at the beginning of a sentence cannot be a mere performative indicator loses force—since that claim is based upon treating "I protest" as having the same use whether used by itself or at the beginning of a longer sentence.[18]

2. To argue that force can be included in meaning, if sense and reference are, seems to me to be like arguing that a man might as well do a second robbery if he has done a first. A wiser alternative might well be that Austin should not have lumped sense and reference together under meaning, but should have followed the more common practice of distinguishing reference from meaning. Anyway, it is unfair of Cohen to suggest that Austin thinks the relation between sense and reference is any closer than that between both of these and force. Austin says, for example,

> I want to distinguish force and meaning in the sense in which meaning is equivalent to sense and reference, just as it has become essential to distinguish sense and reference within meaning.[19]

Further, it is clear from some of Austin's other writings, e.g., *The Meaning of a Word*, that he does not include reference under meaning. Presumably he was willing, in *How To Do Things With Words*, to put together sense and reference, since his main purpose was to distinguish between both of these on the one hand, and other aspects of utterances on the other. This does not mean he thought the lumping-together would be at all advisory for other purposes.

3. It is unclear how we are to take Cohen's claim that certain clarifications which Austin would call illocutionary clarification can properly be called clarifications of meaning. If he is suggesting that the following two clarifications are of exactly the same type, then surely he is wrong:

(a) In saying "He lost his case," I meant he lost his suitcase.
(b) In saying "He is as industrious as my brother," I was praising, not criticizing him.

In (a), I disambiguate a certain ambiguous word. In (b), I explain the point of some possibly quite unambiguous words. If Cohen is merely pointing out that people as a matter of fact say such things as "I meant *S* as praise," or "I mean *S*1 as a command, not a piece of advice," then he is pointing out what every linguist and philosopher also points out at the beginning of a discussion on meaning. Austin himself admits such a sense of "mean," when he says "Admittedly we can use 'meaning' also with reference to illocutionary force—'He meant it as an order,' &c. But . . ."[20] Austin then proceeds, as do other linguists and philosophers, to point out that "mean" and "meaning" are ambiguous. Sometimes, roughly, "meaning" refers to the point of, or the intention behind, an utterance. Sometimes it refers to something quite different—something of course that it is very difficult to spell out, but which might roughly be called the content or topic of the utterance. At any rate, philosophers who point out this ambiguity then usually make it clear that they are not discussing that sense of "meaning" in which "meaning" refers to the point of, or the intention behind, an utterance. Thus, the theories of meaning of, say, Carnap, Quine, Katz, or Ogden and Richards, presuppose that this sense of "meaning" has been excluded. To fit in, then, with the common practice of philosophers and linguists, it is surely wiser and less misleading to use "meaning" in the narrower sense, and to employ other terms, say "illocution," to refer to other aspects of utterances, even though in ordinary language people do use "meaning" to cover a wider field. It is not that difficult, after all, to see that "meaning" is ambiguous. In many languages, for example, we shall not find a single word doing all the jobs that our "meaning" does.

Cohen's objections to Austin's way of talking, then, I find unconvincing. In addition there is a simple, but powerful, reason for preferring Austin's way. If, with Cohen, we make illocutionary force an aspect of meaning then it follows that each time a sentence is used with a different force, it has a different meaning. Any sentence at all, therefore, has multiple (perhaps infinite) ambiguity. However, I know no account of ambiguity ever suggested which would render the sentence "The Second World War was declared in 1939 A.D." ambiguous. The compilers of Webster's or the O.E.D., H. Fowler, the High School student of

---

[18] See M. Furberg, "Meaning and Illocutionary Force," in *Symposium on J. L. Austin, op. cit.*

[19] *Op. cit.*, p. 100.

[20] *Op. cit.*, p. 100.

English, and philosophers and linguists alike would regard ambiguity as a function of the number of semantic readings a sentence can be given, not as a function of the number of intentions with which a man might use a sentence.

As I said before, it does not matter much for our purposes whether we adopt Austin's or Cohen's way of speaking. However, I choose Austin's on the grounds (a) that it will facilitate ease of discussion in the next section, (b) that it is more in keeping with normal philosophical practice, and (c) that it is less misleading anyway.

IV

In this last section, I want to show that thesis (1) is compatible with thesis (3). It might seem difficult to see how this could possibly be so. How can it be the case both that the meaning of a sentence is to be explained with reference to illocutionary acts, and that the meaning of a sentence is always to be distinguished from its illocutionary force, as Austin claims? Further, it may seem clear that if we are to choose between the two theses, it must be thesis (1) that is jettisoned. For surely Austin is right to point out that asking for the meaning of, and the force of, an utterance are quite distinct questions, deserving of quite different answers? That it is thesis (1) that must go appears even more obvious if we put the question in Cohen's terms. It appears that we are forced to reject attempts to explain meaning in terms of illocutionary acts, because it is quite sensible, and may well be true, to say each of the following:

A-1. I know what the sentence meant, but I do not know what illocutionary act the man was performing with it.

A-2. I know what act he was performing, but I do not know what the sentence meant.

A-3. [Particularly relevant to thesis (1).] Sentence *S*1 and *S*2 mean the same, but *S*1 has been used to perform a certain illocutionary act, and *S*2 has not.

Because we can say such things sensibly and truly, we can also say:

B-1. The sentence has a meaning, *and* was used with such-and-such an illocutionary force. (That is, we contrast meaning and force.)

B-2. The sentence could be used to perform that illocutionary act *because* it has such-and-such a meaning. (This surely need not be tautologous.)

If we can say the A- and the B-sentences sensibly and truly, then surely theses (1) and (3) are incompatible, and surely it is thesis (1) that must go overboard.

It seems to me, however, that despite appearances, the theses are not incompatible, and that there is no reason to reject thesis (1) on the basis of the A- and the B-sentences. I shall make my point by way of an illustrative analogy with a quite different philosophical question. I do not want to suggest the analogy can be extended any further than I in fact extend it.

Suppose someone suggests that continuing a mathematical series on a piece of paper is the same thing as knowing how to go with the series. The suggestion would be false. It would be false because we can say each of the following sensibly and truly:

A′-1. He knew how to go on, but did not write anything on paper.

A′-2. He wrote the series down, but he did not know how to go on. It was a fluke.

A′-3. *A* and *B* both knew how to go on, but *A* did, and *B* did not, write the series on a piece of paper.

Because we can say each of the above, we can also say:

B′-1. He knew how to go on, *and* he wrote the series down. (That is, we contrast the two.)

B′-2. He wrote the series down *because* he knew how to go on.

Clearly then, to identify knowing how to go on with continuing the series on a piece of paper would be mistaken. However, for well-known Wittgensteinian reasons, which I shall not repeat here, and which I shall assume to be valid, we do not want to draw the conclusion that the writing and the knowing are two quite distinct activities, running parallel to one another in the way that walking and whistling might. Nor do we want to conclude that the proper analysis of knowing how to go on should contain no reference to activities like writing down the series on paper, or providing verbal answers under questioning, or whatever. We can avoid such conclusions by insisting that the mistake in the original suggestion is not the mistake of equating two distinct activities, but the mistake of equating knowing how to go on with just *one* of an indefinite range of activities that go together to provide the analysis of knowing how to go on. Writing down the series is one, but not the sole, criterion (in appropriate circumstances) of knowing how to go on. Given this, it becomes possible to construe B′-1 and B′-2 in such a way that they do not imply the unwelcome conclusions. In B′-1, the two conjuncts do not refer to two distinct activities. Rather, the first conjunct ("He knew how to go on . . .") refers implicitly to an indefinite range of activities which form the analysis of knowing how to go on, and the second (". . . and he wrote the series down") refers to the particular activity out of that range which, on the occasion, was the criterion for his knowing how to go on. In B′-2, the "because" is not the causal "because." We are not saying that his writing the series down is causally produced by a concurrent series of events forming his knowledge of how to go on. Rather, we explain what he does by pointing out that what he is doing is a member of a range of activities which go to make up the concept of knowing how to go on. It is much more like the "because" in "He applied for the job because he was ambitious." We might put the point crudely by saying: the mistake in the original claim is that of equating the whole with one of its parts. Nevertheless, the relation between the two is one of the whole to a part, and not of two distinct activities. Of course, the part/whole terminology here is metaphorical. Writing down a series is not literally a part of knowing how to go on. Rather, writing the series down is one member out of a range of activities to which reference must be made in explanation of what it is to know how to go on.

I want to suggest that the relationship between the meaning of a sentence and the illocutionary act it is used to perform on an occasion is not unlike that between knowing how to go on and writing the series on a piece of paper. Notice, first, that the A- and the B-sentences are closely similar to the A′- and B′- sentences. That is, the reasons for rejecting an identification between the meaning and the illocutionary act are similar to those for rejecting an identification between knowing how to go on and writing the series on paper. Now it seems to me that, just as rejection of the latter identification does not commit us to regarding the knowing and the writing as two distinct activities, so rejection of the first does not commit us to regarding the meaning and the illocutionary act as two quite distinct aspects of the sentence. And the reasons why we are not so committed are similar in each case. What is wrong with identifying the meaning of a sentence with the illocutionary act performed by it on an occasion is, crudely, that it equates the whole with one of its parts. More exactly, it might well be that the meaning of a sentence can only be explained by reference to a range of illocutionary acts it is capable of performing, of which the particular illocutionary act in question is one member. Given this, we can construe B-1 and B-2 in such a way that they do not commit us to regarding the meaning and the illocutionary act as two distinct aspects of the sentence. In B-1, the first conjunct ("The sentence has a meaning . . .") refers implicitly to an indefinite range of illocutionary acts it is capable of performing, and the second conjunct (". . . and was used with such-and-such an illocutionary force") refers to the particular member out of that range which the sentence was used to perform on the occasion in question. Equally, the "because" in B-2 should not be so construed as to imply that knowledge of the meaning is somehow causally responsible for the quite distinct ability to perform a certain illocutionary act with the sentence. Rather the explanation takes the form of pointing out that the illocutionary act in question is one of a range of acts which we know the sentence to be capable of performing, such knowledge adding up to knowledge of what the sentence means.

It is, of course, true, that the questions "What is the meaning of *S*?" and "What illocutionary act is *S* being used to perform?" are quite distinct questions, demanding different answers. However, this is quite compatible with the idea that meaning is to be analyzed in terms of illocutionary acts. For all it shows is that the first question is to be answered by reference to a wider range of illocutionary acts than will appear in answer to the second question. If Alston is correct, to ask for the meaning of a sentence is to ask for another sentence capable of performing the same range of illocutionary acts. As far as I can see that is quite compatible with Austin's insistence that, on any given occasion, it is necessary to distinguish between the meaning and the illocutionary force. For, of course, it is always correct to insist on a distinction between the wider and the narrower, or the whole and the part, or the general and the particular. I cannot see, then that theses (1) and (3) are incompatible. It will then follow from the previous discussions that each of the four original theses is compatible with every other.

It has not been my aim to pass comment upon Alston's claim that meaning can be analyzed in terms of illocutionary acts; only to argue that the other theses are compatible with that claim. It is worth mentioning, though, that Alston's account of the relationship between meaning and force is certainly no worse than, and indeed is a good deal less opaque than, certain other accounts. To claim, for example, that the meaning of a sentence is an abstract entity such as a proposition does make it hard to see quite what the relationship between meaning and force could be—apart from any other difficulties in such a suggestion.

Still, there are difficulties in Alston's account. I shall briefly mention two of these. First, his account must face all the difficulties which older, "Interchangability *salva veritate*" theories of synonymy meet with. Are "7" and "the only prime between 6 and 10" interchangable in the sentence "He knows that 7 is greater than 5"? If not, can they be synonymous? Well, similarly, does the sentence "He knows that 7 is greater than 5" have the same illocutionary act potential as the sentence "He knows that the only prime between 6 and 10 is greater than 5?" And, if not, can the two mean the same? Clearly the theory must be filled out to deal with such questions.

Secondly, Alston's theory must meet with the problem of "emotive conjugations," with expressions that is which appear to have the same "cognitive" content in Stevenson's sense, but which differ in expressive force. For example "sweat" and "perspire," or "Negro" and "black." Does "He is a black" have the same illocutionary force as "He is a Negro"? If not, does this prevent them from being synonymous, or can we qualify the theory to deal with such a case? Until we can provide answers to such questions, Alston's theory is at best rough and ready. Notice, however, that it is a merit of Alston's approach that it forces us to deal with such problems. More traditional accounts of synonymy, employing the interchangability *salva veritate* test, preclude themselves from dealing with emotive conjugation expressions. For if we concern ourselves solely with the question of whether substitution of "black" for "Negro" affects truth-values, we miss the point about such expressions. Even if they are substitutable *salva veritate*, that should not clinch the question of whether they are synonymous or not. Indeed, it is one of the merits of Alston's approach, as against traditional approaches, that it does not lend unique importance to questions of truth and falsity in settling questions about meaning. For in dealing with illocutionary acts, we are not restricting ourselves to sentences whose prime purpose is to be true. In a way, then, Alston helps fulfil Austin's desire to "play Old Harry" with the "true/false fetish."

University of Miami

Received October 1, 1970

American Philosophical Quarterly
Volume 9, Number 1, January 1972

VIII. ARE TRANSCENDENTAL ARGUMENTS A VERSION OF VERIFICATIONISM?

PETER HACKER

THE increasing employment of transcendental arguments in recent epistemology and metaphysics results partly from the growing understanding of Wittgensteinian doctrines and partly from the general revival of interest in Kant's *Critique of Pure Reason.* The most explicit practitioner is Strawson, in whose writings the Kantian bent is most prominent. Equally important is Shoemaker in whom the Wittgensteinian influence is dominant. Critics of this trend commonly claim that what masquerades as transcendental argument is none other than a version of the verification principle. Ayer[1] levels this accusation against Strawson and Williams[2] intimates as much with respect to Shoemaker. Although few philosophers have pointed out the similarity between Wittgenstein's private language argument and doctrine of criteria and Kantian transcendental arguments, it is no coincidence that these doctrines have been assailed by critics[3] as revivals of verificationism, and misguidedly defended by disciples[4] (not always explicitly) as a new version of the old and tarnished model.

In a recent article[5] Professor B. Stroud gives a cogent general statement of these objections. His conclusion is that

> it looks as if the use of a so-called "transcendental argument" to demonstrate the self-defeating character of scepticism would amount to nothing more and nothing less than an application of some version of the verification principle, and if this is what a Transcendental Argument is then there is nothing special or unique, and certainly nothing new, about this way of attacking scepticism.[6]

It could be, Stroud conjectures, that we are not as far as we might think from Vienna in the 1920s. I suggest that though we have seen through some of the follies of verificationism, we have not yet succeeded in assimilating some of the important truths which, once the Kantian transcendental arguments are stripped of their transcendental idealist implications,[7] are to be found in the *Critique of Pure Reason.*

Stroud sees the sceptic as claiming (a) that a particular class of propositions makes sense, e.g., propositions about objective particulars (or about the past, or other minds, or the future, etc.) and (b) that we can never know whether any of them are true. The contemporary defender of the transcendental type of argument must, in Stroud's view (taking Strawson's arguments in *Individuals* as his paradigm), contend that the truth of (a) is a necessary condition of the meaningfulness of the sceptical doubts expressed in (b) and that the falsity of (b) is a necessary condition of the truth of (a). This refutation of the sceptic, however, though it may be called "transcendental," is in truth a version of verificationism. For it amounts to saying simply that if propositions of a certain kind are meaningful it must be possible to know if they are true or false.

This kind of attack upon transcendental arguments is misguided. First and foremost, it involves a failure to grasp adequately the nature of the sceptical, or more properly—the Problematic Idealist—adversary. Secondly, and consequently, the verificationist reduction involves a deep misunderstanding of the main thrust of the transcendental argument in refutation of the sceptic.

In what follows I shall attempt to give a clearer picture of the position which is adopted by the sceptic. Thereafter I shall touch upon the inadequacy of traditional strategies in response to the sceptic's arguments. To show that transcendental arguments are correct is a task which I shall not attempt here. What I shall try to do is to show the nature of transcendental arguments, where they differ from verificationism, and why, if correct, they succeed in silencing the sceptic while other methods have failed.

---

[1] A. J. Ayer, "The Concept of a Person," in *The Concept of a Person and Other Essays* (New York, 1963), p. 110.

[2] B. A. O. Williams, "Knowledge and Meaning in the Philosophy of Mind," *The Philosophical Review,* vol. 77 (1968), pp. 216–228.

[3] E.g. J. J. Thomson, "Private Languages," *American Philosophical Quarterly,* vol. 1 (1964), pp. 20–31.

[4] E.g. N. Malcolm, *Dreaming* (London, 1959).

[5] B. Stroud, "Transcendental Arguments," *Journal of Philosophy,* vol. 65 (1968), pp. 241–256.

[6] B. Stroud, *ibid.*, p. 255–256.

[7] See P. F. Strawson, *The Bounds of Sense* (England, 1966) where this is so brilliantly done.

I

The sceptic, as this traditional *bête noire* of epistemology appears in modern philosophy, is a direct descendant of Cartesian metaphysics. Many writers on the theory of knowledge have placed the sceptic in the center of the stage in the role of the villain of the philosophical drama. This can be confusing. The sceptic is a straw man, and should not be mistaken for a *malin genie* who, if we cannot lay his spirit, will reduce us to genuine Pyrrhonism. We rightly claim to know of the existence of objective particulars including other persons. We know truths about the past and about the future. How do we justify such cognitive claims? The sceptic's function is to probe the links between premisses and conclusion, between evidence and that which it supports. But if his concern is with the inferential connections, what premisses does he acknowledge, if any? He denies us our conclusion, i.e., our cognitive claim, and maintains that at best we achieve true opinion and not knowledge. But what starting point does he permit? The best way to answer this is to go back to Descartes.

Descartes' sceptical challenger succeeds in throwing doubt upon all cognitive claims concerning objective particulars. But with respect to the contents of Descartes' own mind, his challenge is impotent. For the mind knows with certainty the "thoughts" which occur to it. Cartesian thoughts constitute the foundations of empirical knowledge. I may be mistaken about whether there actually is light that I see, a noise which I hear, heat which I feel. But I cannot be mistaken about my "thought" that it seems to me that I see light, hear noise, feel heat.[8] I know that it seems to me that I see (hear, taste, feel, smell) something or other, simply because it does so seem to me. There is no room for error, just as there is no room, in my subjective account, for evidence. There are many difficulties in the Cartesian doctrine of thoughts.[9] Rather than explore Cartesian exegesis and criticism, I shall proceed to reconstruct a picture of the resultant epistemological theories which stem from the rejection of the Cartesian theodicy (which enables Descartes to bridge the gap between thoughts and statements about objective particulars) together with the retention of the doctrine of immediate knowledge of thoughts.

We justify our cognitive claims about the objective world by reference to evidence. Each piece of evidence produced can be challenged in turn. We can fall back, in our chain of justification, from more exposed positions to less exposed ones. From claims about objective particulars we may retreat to claims about our perceptual experiences of them. From claims about perceptual experiences we fall back on our beliefs about our perceptual experiences (which do not imply that they were experiences of objects.) But there is no further falling back from this position. It is impregnable in that our knowledge of how things seem to us to be does not rest on evidence. It is immediate, or intuitive. How do I know that it seems to me that I see the so and so? Not *any* how—I know because it does seem so to me. Descartes and his sceptical adversary accept this as common ground.[10] The problem which the sceptic now poses is—how can one justifiably infer, from statements about one's "thoughts," statements about objective particulars? Granted that I know that it seems to me that I perceive that a is P, how am I justified in judging that a is P? There is no entailment relation between statements about thoughts of objects and statements about objects. And the arguments from illusion or from dreaming bear witness to the possibility of the truth of the former and falsity of the latter.

The sceptic's stance on other epistemological problems such as our knowledge of the past, of other minds, of universal generalizations, is not in essence different from that described with respect to knowledge of objects. His strategy is simple. He accepts our premisses but denies our conclusions. He acknowledges that we do know the subjective content of our experience, but denies our right to claim knowledge of objects which we allege our experience to be experience of. He sees quite rightly that our sole access to the empirical world is, truistically, through our experience of it. And he wants, reasonably, to know how knowledge of subjective experience can possibly yield knowledge of objective particulars. Put thus, our sceptic is suddenly eminently reasonable. Kant points out very fairly that

---

[8] Descartes—*Meditations* (Haldane and Ross translation, vol. 1, p. 153); see also *Principles of Philosophy* (vol. I, p. 222).

[9] See A. J. P. Kenny, "Cartesian Privacy" in *Wittgenstein; The Philosophical Investigations—A Collection of Critical Essays*, ed. by G. Pitcher (New York, 1966).

[10] As would Kant or Strawson for that matter. Not so Wittgenstein who denies that it makes sense to claim knowledge of one's own immediate experience (*Philosophical Investigations*, §246, pp. 221–224, etc.) His notorious non-cognitive theory of avowals seems to me misguided.

> Problematic idealism . . . which merely pleads incapacity to prove, through immediate experience, any existence except our own, is, insofar as it allows of no decisive judgment until sufficient proof has been found, reasonable and in accordance with a thorough and philosophical mode of thought.[11]

If this crude sketch is, in principle, correct, then it can be seen that Stroud's characterization of the sceptic's position sins by omission. True, the sceptic maintains that propositions about objective particulars make sense, and that we can never know whether or not any of them are true. But he also claims that we can and do know the truth about statements concerning our own "thoughts." It is crucial that *this* claim should not be omitted, for it is upon this point that the whole weight of the transcendental argument against the sceptic, in contrast with the verificationist argument, is brought to bear. Kant wisely points out that the proof required to allay the doubts of the problematic idealist who rejects the Cartesian theodicy must

> Show that we have *experience*, and not merely imagination of outer things; and this, it would seem, cannot be achieved save by proof that even our inner experience, which for Descartes is indubitable, is possible only on the assumption of outer experience.[12]

Before going on to elucidate the nature of a transcendental argument and so to distinguish it from a verificationist one, I shall dwell upon other standard strategies in attempted refutation of the sceptic, to show why, though persuasive, they ultimately fail, and hence why a transcendental argument is vitally necessary.

II

Kant held that there are two kinds of arguments that can be employed in refutation of the sceptic: critical or transcendental arguments and dogmatic arguments. The variety of dogmatic arguments has no doubt increased since Kant's day. Yet it is interesting to note that some of the fundamental objections to them all were made out quite clearly by Kant himself. For he thought that the only way of answering scepticism was by means of some kind of critical argument. Dogmatic arguments are extra-philosophically legitimate, in the sense that they are the inference patterns which, when correctly employed, yield empirical knowledge. But they are incapable of doing the job required of them in the context of the philosophical debate with the sceptic, for they are not self-justifying.

One currently popular method of combating scepticism is to exhibit, with maximum clarity, what our evidential standards are. These standards of reasoning, or of empirical proof, just *are* our ultimate standards by reference to which all our empirical knowledge of the objective world is established. Clearly, there can be no further standard by reference to which our ultimate standards are justified. Hence the sceptic's demand for a further justification amounts to baying at the moon. This method is exemplified in Ayer's discussion of "Descriptive Analysis," an approach which he employs extensively in *The Problem of Knowledge*. Thus, he writes

> We can, for example, show in what conditions we feel confident in attributing certain experiences to others: we can evaluate different types of record: we can distinguish cases in which our memories or perceptions are taken to be reliable from those in which they are not. In short, we can give an account of the procedures which we actually follow. But no justification of these procedures is necessary or possible. One may be called upon to justify a particular conclusion and then one can appeal to the appropriate evidence. . . . There [cannot] be a proof that what we take to be good evidence really is so. And if there cannot be a proof it is not sensible to demand one.[13]

This amounts to agreeing with Kant over the extra-philosophical legitimacy of dogmatic arguments, while claiming wrongly that there is nothing else that can, or needs to, be done.

This dogmatic method of Descriptive Analysis, especially when challenged, slides smoothly into a slightly different one. Our sceptic, having listened patiently to our exposition of our standards of empirical proof in general, replies: "Yes, I can see that this is how we argue and how we think we prove our conclusions. But what makes it reasonable to argue thus, and why is this a satisfactory proof? Since the relation between evidence and conclusion is not deductive (nor for that matter inductive) what makes our procedures rational?" To this challenge the dogmatist replies "Doing this is what 'being reasonable' *means* in such a context."[14] Employing these standards is what we mean by "prove," "confirm," "rational justification." Thus for example Ayer claims that

---

[11] Kant, *Critique of Pure Reason*, translated by N. Kemp-Smith B 275.

[12] Kant, *ibid.*

[13] A. J. Ayer, *Problem of Knowledge* (Baltimore, 1956), pp. 80–81.

> It is characteristic of what is meant by such a sentence as "There is a cigarette case on this table" that my having just the experience that I am having [perceiving-a-cigarette-case-experience] is evidence for the truth of the statement which it expresses.[15]

This dogmatic argument in turn yields a third variant when challenged. For our sceptic's next move is the Humpty Dumpty gambit. Either it is the case that when we use a word it means just what we choose it to mean, or we cannot make words mean so many different things. If the former is true and we have an option, with what justification did we choose to make "rational" mean just this? After all, the sceptic might add, if we are all free and equal, *I* shall mean by "unreasonable" just what you mean by "rational," and *I* shall call "true opinion" what you call "knowledge." The dogmatist must take this alternative, for the other faces him with a mysterious limitation which he is quite unable to explain. Thus he justifies his choice of meanings of "rational," "proof," "reasonable," etc., by the fact that making these words do these particular jobs pays in the hard currency of results. Having the standards that we have cannot be justified by reference to further standards, and giving words these meanings *is* optional, but neither the standards nor the meanings are arbitrary. We have them because they work. They work inasmuch as we find that as a matter of fact, by adopting these standards and using words this way, we are able to discover a wide range of empirical truths. Thus a recent writer claims that our evidential principles of empirical reasoning "[Are] confirmed when it is found that by assuming [them], by inferring in accordance with [them], we can arrive at true beliefs, verified by observation, which we could not have arrived at without assuming the principles."[16] Here we appeal to objective experience to justify the employment of principles of reasoning and evidence which entitle us to claim knowledge of objective experience. To this our sceptic will clearly (and rightly) retort, that it begs the very question at issue. For if we are to justify our standards by reference to their utility in discovering truth, we must already be in a position to validate claims to have established empirical truths about objects on the basis of some evidence or other. But this is precisely what we cannot yet do.

Of course, the dogmatist is not yet exhausted. His next move is to fall back on a conventionalist, coherence theory of Carnapian, or Quinean stamp. To pursue him further at this point would take us into deeper and murkier waters. Rather than do so I shall turn to Kant's own analysis of the dogmatist's argument.

Kant refers to arguments of the kind just adumbrated as "dogmatic" because they amount to little more than an emphatic assertion of our practices. To these assertions the sceptic opposes his equally emphatic denial of the justifiability of our practices. Dogmatic arguments provide no means for settling this conflict. Kant divides all objections to given claims into dogmatic, critical, and sceptical. A dogmatic objection is directed against a proposition, it

> requires an insight into the nature of the object such that we can maintain the opposite of what the proposition has alleged in regard to this object. It is therefore itself dogmatic, claiming acquaintance with the constitution of the object fuller than that of the counter assertion.[17]

It is precisely this that is wrong with it. For the sceptic claims no lesser insight and denies the dogmatist's conclusions. This alone suffices for a sceptical victory over conventional dogmatism.[18] If the argument is conducted on the dogmatist's grounds, the sceptic wins if he does not lose.

The critical philosophy is opposed to dogmatic arguments of the second kind outlined above because they assume that progress can be made and philosophical justification demonstrated according to principles, from concepts alone. Arguments of this kind, as all paradigm case arguments, assume

> that it is possible to do this without first investigating in what way and by what *right* reason has come into

---

[14] P. F. Strawson, *Introduction to Logical Theory* (England, 1952), p. 257.

[15] A. J. Ayer, *Problem of Knowledge*, *op. cit.*, pp. 132–133.

[16] H. Whiteley, "Epistemological Strategies," *Mind*, vol. 78 (1969), p. 32.

[17] Kant, *op. cit.*, A 388.

[18] Kant, *op. cit.*, B 23.

> possession of these concepts. Dogmatism is thus the dogmatic procedure of pure reason without previous criticism of its own powers.[19]

It is not that we should deny that these are our standards; clearly they are. Nor is it that we mean something different by "rational," or "confirm" or "proof." Moreover it is true that these standards yield empirical knowledge. The trouble is that we cannot show the latter proposition to be true by dogmatical argument without begging the question, and we cannot demonstrate a justification of our standards and our word meanings, as opposed to the mere fact that these *are* our standards, without a critical non-dogmatic argument. The recourse to experience itself in the third dogmatic ploy not only fails by being viciously circular, it involves a failure to understand where the problem arises, and hence where it must be solved. Thus Kant emphasizes that

> We have no right to ignore these problems, as if their solution really depended upon the nature of things. . . . For since reason is the sole begetter of these ideas, it is under an obligation to give an account of their validity or of their illusory dialectical nature.[20]

The problem is that of justifying those modes of reasoning employed in the establishment of empirical truths. The unfolding of experience itself is at best an illustration of these principles, and not a justification of them. No appeal to experiential knowledge can solve this problem, for the principles in question are themselves required to establish any experiential knowledge that we might wish to appeal to. If we do know that these standards of proof and justifications of cognitive claims are adequate, then this knowledge is *a priori*. But to show that we do know this,

> the analysis of concepts is useless, since it merely shows what is contained in these concepts, not how we arrive at them *a priori*. A solution of this latter problem is required that we may be able to determine the valid employment of such concepts in regard to the objects of all knowledge in general.[21]

Such a solution is provided by the Critical Philosophy in general, and by transcendental arguments in particular.

III

Kant labored hard to elucidate his technical term "transcendental." Briefly transcendental knowledge amounts to knowledge of the necessary conditions of conceptualized experience. The discovered necessities are ascribed to the requirements of receptivity and spontaneity, and limited to discursive intelligences. Transcendental knowledge is prior to experience in the sense that it is knowledge of those conditions in virtue of which experience itself (i.e., knowledge of objects of sense perception) is possible. In one of many attempts at defining transcendental knowledge Kant suggests that it is "all knowledge which is occupied not so much with objects as with the mode of our knowledge of objects insofar as this mode of knowledge is to be possible *a priori*."[22]

We have seen that the sceptical problem is to be conceived of as a challenge to justify our claim to know *a priori* that knowledge of subjective experience under certain specifiable conditions legitimizes claims to knowledge of objective particulars. The claim that such knowledge is *a posteriori*, which is the third dogmatic ploy, is incorrect. The equally dogmatic claim that such transcendental knowledge can be established by arguing from mere concepts is explicitly denied by Kant. No objectively valid synthetic judgment can be derived directly from concepts. Rather,

> Through concepts of understanding pure reason does, indeed, establish secure principles, not however directly from concepts alone, but always only indirectly through relation of these concepts to something altogether contingent, namely *possible experience*.[23]

It is not necessary, for the purpose of showing how different transcendental arguments are from verificationism, to demonstrate the general validity of Kant's Transcendental Deduction. It suffices for present purposes to suggest its main drift. Kant's central achievement was not, as he thought, the Copernican Revolution with its transcendental idealist implications. It was his grasp of the necessary unity of apperception, its significance and its implications. Both the Transcendental Deduction and the Refutation of Idealism turn on the same point, namely that inner experience is only possible if experience of objects is possible too. It is a contingent fact that we exist and have any perceptual experience at all. But given that we do, and given that we can only have empirical knowledge by applying concepts to given intuitions, then, Kant claims, the argument desired to establish the general conclusion can get off the ground.

---

[19] Kant, *op. cit.*, B xxxv (my italics).

[20] Kant, *op. cit.*, A 763/B791.

[21] Kant, *op. cit.*, A 94/B126.

[22] Kant, *op. cit.*, A 12/B25, see also A 56/B80, A 85/B117, etc.

[23] Kant, *op. cit.*, A 737/B765.

The Deduction and Refutation jointly provide the core of Kant's general answer to the sceptic. His specific answers to scepticism about causation, induction, and substance come in the Analogies. But the central reply concerns the mutual dependence of knowledge of the mind and knowledge of objects. The Deduction is more detailed and abstract, and its arguments are conducted in terms of the necessary conditions of transcendental self-consciousness. It elaborates the implications of the original unity of apperception, stressing the requirement of the possibility of distinguishing how (phenomenal) objects are subjectively judged to be from how they are. This is necessitated by the dual requirements of the unity of a thought and of original self-consciousness or self-ascribability of experience, without which cognitive experience is not possible. The Refutation is more compressed, and its argument is conducted solely in terms of one of the necessary conditions of *empirical* self-consciousness, namely the possibility of self-ascription of temporarily successive experiences *as* temporally successive. This condition, Kant suggests, is uniquely satisfied by a series of experiences which, for the most part, can be taken to be experiences of (relatively) permanent objects whose existence is independent of being perceived.

What then is the relationship between statements about subjective experience and statements about objective particulars? The sceptic claimed that we cannot justify our claims to know the latter statements on the basis of the former. We should concede that the relationship cannot be one of entailment. Moreover we should deny that the connection can be inductive and contingent, not simply because this will concede the whole case to the sceptic, nor simply because principles of inductive reasoning are themselves in question in this context, but because it will grant that the Empirical Idealist's picture of the world is a picture of possible experience. But this Kant has shown to be false. Contrary to the sceptic's allegations we can validly distinguish between (phenomenal) reality and illusion. We do so by reference to those regularities of coexistence and succession embodied in the most general way in the concepts of substance, causation, and reciprocity which Kant valiantly struggles to elucidate and justify in the Analogies. Those perceptual experiences which do not cohere in the requisite rule governed way with the bulk of our experiences are rejected as objectively invalid. There is nothing novel in this nor does Kant suggest that there is. There *are* the standards by which we distinguish the illusory from the real. What is novel is the justification of our commonplace standards as resting on a more profound requirement of reason than convention or pragmatic success.

Yet although this sketch of an argument goes a long way, it does not render a completely satisfactory answer to the previous question. If the nexus is neither entailment nor empirical, what is it? The answer to this point is provided by the contemporary Wittgensteinian doctrine of a criterion, as that which is necessarily good evidence for a given thing The combination of this doctrine with the core of Kantian transcendental arguments, stripped of their idealist trappings and of the complexities of transcendental psychology, provides an answer to the sceptic, which, if correct, is far superior to the customary dogmatic strategies so popular at the moment. The notion of a criterion is the most important as well as the least understood part of the Wittgensteinian legacy.

No doubt ordinary language philosophers, supporters of paradigm case arguments, various brands of conventionalists or positivists, and many others will deny the correctness of the general drift of transcendental arguments which has been crudely sketched here, rather than argued. It is not my purpose in this paper to provide the demonstration of the validity of some version of a transcendental argument. Rather it is to show objectors where their proper target lies. It does not lie with either crude or sophisticated versions of verificationism. What must be shown is *either that original apperception is not a necessary condition of experience but is, in some way, a logical construction, or that the possibility of self-ascription of experience does not require, as one of its necessary conditions, the possibility of knowledge of objects.* Conversely, if the sceptic is to be refuted it can only be by an argument of the kind outlined. The crucial question to be answered, as both Kant and Wittgenstein realized, is what are the necessary condition of criterionless or original (underived) self-ascription of experience? It matters not whether the answer concentrates upon the conditions for self-ascription of experience of (phenomenal) objective particulars, as in Kant, or upon the conditions of self-ascription of "private objects" such as pain, as in Wittgenstein. For the two are inevitably linked.

If the general line of argument is correct then we may claim with Kant, in the Refutation and Paralogisms, that our experience of objects is immediate, and not through a diaphanous veil of ideas. Although our only access to the objective world is *via* our experiences of it, our experiences do provide us with sufficient grounds for knowledge of items other than the experiences themselves, namely the independently existing objects of which they are experiences. Our judgments about objects thus experienced are (normally) non-inferential. Our justifications of such judgments are inferential, but criterial and not inductive. Our ultimate criteria for our judgments about objective particulars are our subjective judgments of our perceptual experiences. These we know directly or non-evidentially. But this in turn is only possible if these judgments are necessarily good evidence for judgments about objects. This provides the required room for occasional illusion, while blocking off both the sceptic and the phenomenalist. Consequently the price of both dogmatic idealism and of scepticism is silence.

IV

We are now in a position to return to our initial discussion to see how a transcendental argument differs from a verificationist one. The verificationist does indeed argue that the falsity of the sceptic's claim that it is impossible in principle to know whether or not statements about objective particulars are true or false is a necessary condition of the meaningfulness of such statements, and hence a necessary condition of the claim making sense. But the sceptic might accept this with equanimity. Of course statements about objective particulars are hypotheses "which transcend all possible experience. How could such a hypothesis be backed by meaning?"[24] Far from refuting scepticism, it reinforces it. The verificationist specification of the conditions of meaningfulness of a restricted class of statements leaves it open to the sceptic to reject the whole class while retaining intact the privileged statements about subjective experience. Furthermore it may appear that transcendental arguments yield only conditional necessities of our present ways of thought.[25] But a radical conventionalist may claim that we may adopt different ways of thought for which these alleged necessities do not hold.

Three questions then must be answered if the case against the assimilation of transcendental agreements to verificationism is to be completed. (1) How does the transcendental argument against the sceptic differ from the verificationist one? (2) Can the sceptic accommodate the transcendental argument with the equanimity with which he accepts the verificationist one? (3) Does the conditionality of the transcendental argument which draws a conceptual connection between inner and outer experience leave room for a conventionalist riposte?

We need not follow the verificationist in claiming simply that the falsity of the sceptic's claims is a necessary condition of their meaningfulness. What we should suggest is that if the sceptic's claim is true, then his other claims, namely to know how things seem to him to be, are thereby made meaningless. The sceptic can deny the possibility of knowledge of objects, but he cannot both affirm that he knows the nature of his inner experience and *simultaneously* deny that his knowledge of his inner experience is ever sufficiently good evidence to enable him to know how things actually are. A necessary condition for the meaningfulness of "I know that it seems to me that I perceive that *a* is *P*" is precisely that this proposition be necessarily good evidence for *a*'s being *P*. Bringing one's subjective perceptual experience under general concepts requires that the experiences thus conceptualized constitute adequate grounds for the application of those general concepts to independently existing particulars conceived of as the objects of these experiences. The sceptic could accept the verificationist's argument and deny the meaningfulness of statements about objective particulars. The verificationist has grasped the wrong end of the stick, leaving the sceptic securely in possession of his array of "Cartesian thoughts," out of which he may, so he conjectures, build up a Humean *fiction* of an objective world. But he cannot meet the transcendental argument in the same way. For it allows him to retain his position only at the expense of denying the sense of statements about his inner experience. And this, as we have seen, not even the sceptic will do.

What, finally, of the accusation that the necessities revealed by transcendental arguments are merely conditional, that they merely reveal the necessary conditions for thinking as we do, and hence that we could adopt different ways of thought for which these necessities do not hold? Certainly there is a sense in which the alleged necessities are conditional, but it is not that they are simply conditions of a purely conventional and hence alterable way of thought. Their conditionality consists in the fact that the occurrence of any cognitive experience at all is itself contingent and not necessary. But cognitive experience is intuition conceptualized in a certain way. Could our intuitions not be conceptualized in relevantly different ways, such that the necessary link between the inner and the outer did not hold? After all, if, as I have suggested, the Wittgensteinian doctrine of criteria must be added to Kant's theory of the relations between the subjective and objective judgment of experience, and if criteria are, as Wittgenstein suggests "fixed by a definition" why should we not define things otherwise? It appears that Wittgenstein's theory reintroduces a conventionalism which Kant labored to exclude? Such an appearance is however illusory.

---

[24] Wittgenstein, *The Blue and Brown Books* (England, 1958), p. 48. The same line of argument is skillfully developed by Stroud—*op. cit.*, p. 251.

[25] See Stroud, *op. cit.*, p. 252.

One must here distinguish two moves. The common or garden conventionalist may claim that differentiating our perceptual experience by means of concepts of material objects occupying a single unified spatio-temporal system is convenient but dispensable. There are other possible ways of constructing a system of concepts of objects, of which Whitehead's conception of basic particulars as slices of continua of four dimensional events may be an example.[26] But such suggestions apply to the mode of characterizing the "object" end of the subject-object relation. Both Kant and Strawson would object to such latitude in the construction of possible categorial schemes. Whether they are right in stressing the necessity of differentiating experience in the way we do or not, is not to the point. The issue is that in the controversy with the sceptic who denies that inner experience is ever good evidence justifying knowledge of objects, the conventionalist must frame his criticism of the Neo-Kantian in a quite different way. For if, *ex hypothesi*, he concedes that the alleged necessities *are* necessary to our conceptual scheme, and if he concedes, as I suggest he should, that the necessity is a condition of the possibility of the unity of apperception, or, less abstractly, of self-ascription of temporally successive experiences as temporally successive, then the only option left is the suggestion that there can be a kind of experience which need not be conceived of as the experience of a subject. For the link between subjective and objective judgment of experience is forged prior to the resolution of the question of the specific nature of the concepts of objects which experience is conceived to be experience of. Hence if the conventionalist is to challenge the transcendental argument by saying that the revealed necessities are only conditional upon our thinking of the world in the way we do, it must be remembered that the challenge is only legitimately directed at the "*subject*" end of the subject-object relation. The necessities are conditional upon our thinking of our own subjective experience in the way we do, namely as ours, and as known underivatively. The mind boggles at an alternative to this. For the conventionalist must now claim that it is possible to imagine a philosophy that is diametrically opposite to solipsism. He must eliminate the notion of the self, and with it self-consciousness, and conceive of all consciousness as distributed over the world as a whole.[27] This is no easy task.

Stroud concluded that we need to know whether or not some version of verificationism is true, and that this is what must be discussed by those who favor the Kantian turn in recent philosophy. I suggest that we have travelled a long way since Vienna in the 1920s. What we do need to know is whether some version of the Transcendental Deduction, and the transcendental argument to establish a necessary link between inner and outer is true. This is a very different requirement. Until it is satisfied scepticism will not be refuted. Paradoxically, we are not so far as we might think from Koenigsberg in the 1780s.

St. John's College, Oxford

Received October 28, 1970

---

[26] See S. Körner, "Transcendental Tendencies in Recent Philosophy," *Journal of Philosophy*, vol. 63 (1966), p. 559ff.

[27] Wittgenstein, "Notes for Lectures on 'Private Experience' and 'Sense Data'," ed. by R. Rhees, *The Philosophical Review*, vol. 77 (1968), p. 282.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 1, January 1972

IX. INDIFFERENCE AND MORAL ACCEPTANCE

M. B. E. SMITH

I SHALL here examine the contention that no one can accept a moral judgment unless he has a favorable or an unfavorable attitude toward that thing which the judgment is about. Borrowing the terms from W. D. Falk.[1] I shall call this view of moral acceptance "internalism," and those who accept it, either explicitly or by implication, "internalists." Internalism, as I have defined it, has rarely been avowed explicitly,[2] but it is implicit in positions taken by many recent philosophers. It is, quite obviously, implied by C. L. Stevenson's contention that two persons can reach moral agreement only if they share a favorable or an unfavorable attitude.[3] It is also implied by the emotivist contention that all utterances which express moral judgments express favorable or unfavorable attitudes.[4] For, suppose that internalism is false, that someone can accept the moral judgment that, e.g., segregation is morally wrong, and yet lack an unfavorable attitude toward segregation. Then, if he utters the sentence "Segregation is morally wrong," his utterance plainly expresses a moral judgment, even though it does not express an unfavorable attitude. Surely, a sufficient condition of an utterance's expressing a moral judgment is that it expresses the moral opinion of its speaker. Finally, internalism is implied by the prescriptivist contention that accepting a moral judgment requires having some inclination to act in accordance with it.[5] For, to be inclined to act in accordance with the judgment that segregation is morally wrong is *inter alia* to be inclined not to practice segregation, to do whatever one can to persuade others not to do so, and to be disposed to feel at least twinges of remorse, shame, or guilt if one fails to act in these ways. And, that a person has such inclinations and dispositions and is able to avow sincerely that segregation is morally wrong is surely a sufficient condition of his having an unfavorable attitude toward segregation. Hence, if accepting a moral judgment requires having an inclination to act in accordance with it, then, since moral acceptance also requires that one can sincerely assert a sentence which expresses the judgment, it follows that moral acceptance requires that one have a favorable or an unfavorable attitude toward that which the judgment is about.

---

[1] Falk introduced the term "internalist," together with the term "externalist," which I shall also use, in his paper " 'Ought' and Motivation," reprinted in J. Hospers and W. Sellars, (ed.), *Readings in Ethical Theory* (New York, 1952). He introduced these in discussing whether judgments of obligation are internally related to motivation in such a way that the person who thinks he has an obligation necessarily has some motivation toward fulfilling it. Since to be motivated to fulfil one's obligation is to have a favorable attitude toward fulfilling it, my use of these terms differs from Falk's primarily in that I do not confine them to contentions restricted to judgments of obligation.

[2] In fact, I have been unable to discover any philosopher who explicitly adheres to the theory as I have just stated it, but P. F. Nowell-Smith and Gilbert Ryle have come close. Nowell-Smith holds that moral principles are a kind of "pro-attitude" in *Ethics* (Harmondsworth, 1954), p. 306f. Ryle has contended that no one can know that something wrong has been done without disapproving of it or being ashamed of it in "On Forgetting the Difference Between Right and Wrong," in *Essays in Moral Philosophy*, ed. by A. I. Melden (Seattle, 1958), p. 152.

[3] This contention appears at many places in Stevenson's writings. But see in particular: *Facts and Values* (New Haven, 1963), p. 1f; and *Ethics and Language* (New Haven, 1944), pp. 1–19, 30–31, 134f.

[4] It would be tedious to list all of the philosophers who have advanced this emotivist contention, but a partial listing would include: A. J. Ayer, *Language, Truth and Logic*, 2nd ed. (London, 1946), p. 105f.; C. L. Stevenson, *Facts and Values* (New Haven, 1963), p. 15f.; *Ethics and Language* (New Haven, 1960), p. 13f.; and Paul Edwards, *The Logic of Moral Discourse* (New York, 1955), in particular p. 147f.

[5] Cf., R. M. Hare, *The Language of Morals* (Oxford, 1951), ch. X, pp. 163–179. Hare argues that in order for someone to accept a moral judgment in the "evaluative" sense he must accept some "imperative" derivable from it (p. 171). Hare's position is, therefore, seemingly more complicated than the above, but it is not really. In the first place, it is plain that Hare thinks that accepting a moral judgment in the "evaluative" sense is really accepting a moral judgment in the common-garden sense; for he holds that, if someone accepts a judgment in a "non-evaluative" sense, he accepts it only in an "inverted-commas" sense, i.e., accepts that it expresses the moral opinion of some group of persons which does not include the speaker (p. 164f.). Secondly, according to Hare, if I accept an "imperative" derivable from a moral judgment, I must either have some inclination to act in accordance with it, or, at the very least, feel guilt if I fail to so act (pp. 164–165, 169). But, to intend to act in accordance with a moral judgment or to feel guilty if I fail to act in accordance with it is to be inclined to act in accordance with it. Hence, Hare's position is equivalent to the view that accepting a moral judgment requires having some inclination to act in accordance with it.

My object in this paper is to show that the argument in favor of internalism is quite inconclusive, and that the evidence now at hand gives us as good reason to accept the opposing position. Again borrowing from Falk, I shall call this opposing position "externalism," and those who accept it, either explicitly or by implication, "externalists." Externalism holds, in brief, that a person may accept a moral judgment about a thing even though he lacks a favorable or an unfavorable attitude toward it—or, as I shall say, even though he is "indifferent" toward the thing. In contrast to internalism, externalism is not widely held. I know of but three contemporary philosophers who may reasonably be thought externalists. Philippa Foot[6] is unmistakably one, and William K. Frankena[7] and P. T. Geach[8] appear to lean toward externalism.

I

Internalists have apparently thought their position self-evident, or virtually so, for they have not bothered to argue strenuously for it.[9] Such arguments as they have offered are all, so far as I can determine, of the following form. Internalists first set out a case in which a person utters a sentence which expresses some moral judgment and then claims to have no inclination to act in accordance with the judgment or claims not to have the favorable or unfavorable attitude which is usually expressed by utterance of the sentence. They then invite the reader to see that such a person could not have expressed a moral opinion. Usually, they suggest that the speaker's utterance cannot be taken literally and that he must be interpreted as saying something like "Most persons accept this judgment, but I don't."[10] Some internalists, in addition, argue that the speaker's utterance cannot be taken literally because his added profession of indifference shows that he fails to understand the meaning of the sentence he uttered.[11]

Such arguments give a difficult question much too summary treatment. Suppose that we are arguing with a man who maintains that there is nothing wrong in racial segregation. Our argument is over such considerations as these: he maintains that, with very rare exception, black people are far less intelligent than white people, that they are far more prone to immoral behavior than the most degenerate whites, and finally, that if they are not discriminated against, both legally and socially, the races will inevitably commingle with disastrous results. We argue with him as best we can, perhaps citing several anthropological and sociological studies. Finally, he says, "All right, you've convinced me. Segregation is morally wrong." But, when we suggest that he act in accordance with the moral judgment he now professes to accept, perhaps by contributing to some such organization as the NAACP, he replies, "Certainly not. You've convinced me that segregation is morally wrong, but what of that? I care about things only when they affect my interests; and, since racial segregation neither benefits me nor harms me, I am completely indifferent as to whether the practice continues."

How should we interpret this man's utterance? In most such instances we should probably believe that in saying "Segregation is morally wrong" he had not been sincere. We should think that our argument had failed to convince him and that, in pretending to agree with us and then professing indifference toward segregation, he had played a childish trick upon us. But this interpretation of his utterance is uninteresting, given our present purposes, for it supports neither internalism nor externalism. Since on this interpretation we do not take him to be sincere in saying that segregation is morally wrong, the question does not arise as to whether he can believe it to be wrong and yet remain indifferent toward it. The present example is relevant to the dispute between internalism and externalism only if it be presumed that the speaker intended to state what he believed to be his true moral conviction.

---

[6] "Introduction," *Theories of Ethics*, ed. by Philippa Foot (Oxford, 1967), p. 9f.

[7] "Recent Conceptions of Morality," in *Morality and the Language of Conduct*, ed. by H. N. Castaneda and G. Nakhnikian (Detroit, 1963), p. 20.

[8] "Good and Evil," *Analysis*, vol. 18 (1957–1959), p. 39f.

[9] In contrast, there have been a large number of arguments offered in favor of restricted internalism—that confined to judgments of obligation. Fortunately, I need not consider them here. William Frankena has canvassed all such arguments and has shown decisively that none is successful in "Obligation and Motivation in Recent Moral Philosophy" in *Essays in Moral Philosophy*, *op. cit.*

[10] See: Stevenson, *Ethics and Language*, *op. cit.*, p. 16.; Hare, *The Language of Morals*, *op. cit.*, p. 164; Ryle, "On Forgetting the Difference Between Right and Wrong," *op. cit.*, p. 152.

[11] Hare, *The Language of Morals*, *op. cit.*, p. 171f. For similar arguments directed toward related positions, see: G. E. Huges, "Moral Condemnation," *Essays in Moral Philosophy*, *op. cit.*, p. 116; A. Moore, "A Categorical Imperative?," *Ethics*, vol. 58 (1952–1953), p. 237f.; and John Ladd, "The Desire to Do One's Duty," *Morality and the Language of Conduct*, *op. cit.*, p. 423f.

If it be stipulated that the man spoke sincerely, the internalist will then doubtless contend that, despite the speaker's intention, his utterance cannot be taken literally. He will perhaps argue that we cannot interpret it literally because his profession of indifference shows that he does not understand the meaning of the sentence "Segregation is morally wrong." Since he does not understand the sentence's meaning, what he must have meant in asserting it is something like "Speakers of English who believe what I now believe about the characteristics of black people would say that segregation is morally wrong." Indeed, the internalist must claim that the speaker's grasp of the sentence's meaning is defective: internalism collapses if this is not so. If he admits that a person, who is indifferent toward segregation and has sincerely asserted the sentence "Segregation is morally wrong," can yet fully understand the sentence's meaning, the internalist can hardly deny that this person accepts the moral judgment expressed by the sentence. Surely, a sufficient condition of accepting a moral judgment is that one fully understand and can sincerely assert some such sentence as this.

The contention that such a person cannot fully understand the meaning of the sentence "Segregation is morally wrong" is, without doubt, intuitively plausible. But, there is good reason to reject it. It must be remembered that, in the present example, the speaker has given cogent reasons against the judgment that segregation is morally wrong. The propositions offered as reasons are doubtless false, but they are certainly relevant to the acceptability of the disputed judgment. Were it true that black people are inherently inferior, prone to immorality, etc., most of us would then consider there to be good reason (although perhaps not conclusive reason) to think racial segregation morally permissible. Now, the following is surely a reasonable principle of argument: if we are arguing with someone and he advances what we take to be relevant opposing reasons to the position we maintain, we then have at least *prima facia* reason to believe that he understands the meaning of the sentence which expresses the disputed proposition or judgment. Therefore, given this principle, there is good reason to suppose that the speaker in the example above does understand the meaning of "Segregation is morally wrong." Of course, that he can carry on a moral argument does not entail that he understands the sentence's meaning; and so the example cannot be regarded as a refutation of internalism. The internalist may contend that his proficiency in moral argument shows only that he has learned how we apply our moral vocabulary without ever really understanding it. Still, it is undeniable that the indifferent man's conduct of the argument is excellent reason to think that he understands the sentence's meaning. And, this alone shows that the burden of proof is on the internalist: he cannot just claim that anyone who sincerely asserts the sentence "Segregation is morally wrong" and is indifferent toward segregation thereby shows that he does not fully understand the meaning of the sentence he uttered.

Faced with this challenge the internalist might appeal to a non-cognitive analysis of the meaning of "morally wrong," arguing: "It is simply part of the meaning of such expressions as 'morally wrong' that the person who sincerely asserts a sentence of the form '*X* is morally wrong' cannot be indifferent toward *X*, just as it is part of the meaning of the word 'nigger' that the person who sincerely asserts of a black man that he is a nigger must have an unfavorable attitude toward him."

But this advances the internalist's cause not a whit. For, how is the internalist to support his contention that "morally wrong" has some noncognitive meaning? There would seem to be three alternatives. First, he might accept a Stevensonian account of meaning,[12] holding that "morally wrong" counts as having emotive meaning in that it has a disposition to express and to evoke unfavorable attitudes. But, that "morally wrong" has emotive meaning in this sense is not sufficient to give the internalist what he wants. The word "communist," for example, has in this country a disposition to express and to evoke unfavorable attitudes, but no one would contend that it is part of the meaning of "communist" that anyone who sincerely asserts "Mao is a communist" must have an unfavorable attitude toward Mao. Secondly, relying upon the "meaning as use" dogma, the internalist might contend that "morally wrong" has noncognitive meaning in the sense that someone misuses the expression if he sincerely asserts "*X* is morally wrong" and is indifferent toward *X*. This, however, is hardly any conclusive argument. The externalist, if he accepts the "meaning as use" dogma, will simply deny that such a use of the expression "morally wrong" is a misuse of it; and there is certainly no generally accepted theory of what constitutes a misuse of language to which the internalist may appeal to confute him. Finally, the internalist might contend that "morally wrong" is shown to have emotive meaning by the fact that anyone who sincerely asserts "*X* is morally wrong" and is indifferent toward *X* cannot count as understanding the full meaning of "morally wrong." But, in the present connection this last argument is obviously question begging.

---

[12] *Ethics and Language*, *op. cit.*, Chap. III.

So far as I can determine, no internalist has ever recognized that there can be, in some cases, good reason to suppose that a person who says sincerely that something is wrong and is indifferent toward it can yet count as understanding the meaning of the sentence he uttered. Perhaps this is why internalists have been content merely to assert that such a person's knowledge of the meaning of moral language must be imperfect. But, until some internalist gives us a compelling argument for this contention, cases like that of the man who is indifferent toward segregation can hardly be thought to be conclusive arguments against externalism.

II

Although the main internalist argument is inconclusive, there are certain features of our ordinary moral discourse that appear to support their position strongly. First, when someone asserts a moral proposition we ordinarily expect that his attitudes are engaged. If someone tells us that segregation is wrong we do not expect him to be indifferent toward the practice of segregation. Were someone to say "I think racial segregation to be absolutely iniquitous, but I do not care whether the practice is overthrown," his utterance would have a paradoxical flavor, one which Stevenson suggests is akin to that of "The cat is on the mat, but I don't believe it."[13] Secondly, our standard of success in moral argument is not a cool, indifferent agreement, but rather the adoption of certain attitudes and behavior. Were we to set out to persuade a segregationist of the evils of his practice, we should feel that we had succeeded only if we remarked some change in his behavior toward black people. Merely bringing him to say "Oh, yes, segregation is wrong—interesting point that!" would not satisfy us. Thirdly, if someone says that something is morally wrong and we know that he is indifferent toward it, we certainly would regard his indifference as good reason to think him insincere. Finally, moral disagreement normally involves disagreement in attitude. If two persons disagree on whether segregation is wrong, we expect that they will have opposing attitudes toward a number of things. We should expect that things which might make the one downhearted, fearful, or angry would tend to encourage or please the other. If they were equally unmoved by events bearing upon segregation, e.g., a Supreme Court decision that strikes a blow against the practice, we might well wonder what their disagreement came to.

At first glance, these four points make internalism seem extremely plausible. Nevertheless, they certainly do not establish its truth. What they establish is that moral judgments are nearly always expressed in an atmosphere of moral concern; and so, when we are concerned to discover what moral judgments a man accepts we have a very great interest in determining his attitudes. This is surely not incompatible with externalism. Still, it may be thought that this concern with attitudes is, to paraphrase Mill, a consideration "capable of determining the intellect" to accept internalism. It may be urged that this concern surely marks off moral judgments from straightforwardly factual ones. When we wish to know what factual propositions someone accepts we have no great interest in his attitudes. If we want to know whether someone thinks that Ferraris are faster than Volkswagens we might arrange a race and see which he bets on, but we should not think it necessary to canvass his attitudes.

However, it would seem that the externalist can explain this difference between moral and nonmoral judgments without compromising his position. He may argue, "Everyone who is seriously concerned with moral matters is more concerned with men's actions than with their beliefs. In situations that make moral demands upon persons we are not concerned with what they believe about their respective situations except insofar as this has a bearing upon what they will do. And so, when the moral convictions of others are important to us, we are vitally concerned with their attitudes. A man may hold correct moral beliefs about a situation, but we cannot in general expect that he will refrain from the wrong unless we have some assurance that he disapproves of wrongdoing. And so, were someone to tell us that he has a certain moral belief and then confess that he has no inclination to act upon it, his utterance would be puzzling to us. For he has at once told us that he has a certain moral belief and ensured that his information is irrelevant to our usual concern in discovering his moral conviction. But, that his utterance is puzzling in no way shows that he does not hold the moral belief he ascribed to himself. Furthermore, from experience we have learned that certain moral beliefs are virtually always accompanied by certain attitudes. Those who think something morally wrong are almost always against it, and those who believe that they have an obligation to do a thing virtually always have at least some inclination to do it. And so, a person's failure to have the appropriate attitude is good inductive evidence that he does not hold a certain moral belief. In contrast, when we are concerned to discover a man's straightforwardly factual beliefs we often do not care how or whether he will act on his beliefs. Nor are most such beliefs accompanied by specific attitudes. We have not learned from experience to expect that a particular attitude will accompany the belief that Ferraris are faster than Volkswagens."

---

[13] *Facts and Values*, *op. cit.*, p. 207.

The externalist can, then, give a persuasive explanation of why we are interested in determining a man's attitudes when we wish to determine his moral convictions; and so he is not forced, by recognition of the four features of moral discourse cited earlier, to concede that internalism is a more plausible position than his own. Of course, his explanation will hardly be palatable to an internalist, who perhaps may be tempted to respond by saying "The explanation supposes that the possession of attitudes is only contingently related to moral belief. But, there is a necessary connection here: that a person lacks an unfavorable attitude toward segregation is not merely good evidence that he does not think it wrong; it establishes that he cannot think this." But this would be merely to assert internalism, not to argue for it.

III

Still, it may fairly be said that internalism has a strong intuitive appeal—much more so than does externalism. The burden of argument, therefore, rests with the externalist. If he wishes to show that his position is no less plausible than the internalist's, he must do more than to counter his opponent's arguments: he must, in addition, explain why internalism has seemed virtually self-evident to so many philosophers. I shall now try to provide him with such an explanation. To do this, I must first distinguish between two kinds of acceptance and agreement.

There is, in ordinary usage, a sense of the word "assent" in which to assent to what is expressed by an utterance is to adopt the attitude expressed by the utterance. This form of acceptance is, I think, most common with regard to utterances of those sentences that are very frequently used to express some particular attitude or set of attitudes. "Jones eats like a pig" is such a sentence, being a conventional means of expressing disgust. Let us suppose that Brown utters the sentence to Robinson, thereby expressing disgust. If Robinson believes it appropriate to say that Jones eats like a pig and if he shares Brown's disgust, it would be perfectly natural—albeit perhaps somewhat formal—to say that Robinson assented to what Brown had said. But, suppose that Robinson is not disgusted and that he even approves of Jones's table manners, thinking them "natural." We should then say, I think, that Robinson did not assent to what Brown had said, even though Robinson may believe that, given the way the expression "eats like a pig" is ordinarily applied, it certainly is not wrong to say that Jones eats like one.

In this sense of the word, assenting is a form of acceptance other than mere belief: it requires not merely that the proposition or judgment expressed by what someone says be accepted, but also that the expressed attitude be shared. Let us return to the case in which Robinson does not assent to Brown's utterance of "Jones eats like a pig," because he approves of Jones's table manners. It would be perfectly natural to say of this case that Robinson does not accept what Brown had said about Jones.[14] Moreover, a lack of assent can, in some cases, be a form of disagreement. Because Robinson does not share the attitude expressed by what Brown said, we may say that Robinson disagrees with what Brown said about Jones. Indeed, were Robinson to affect to agree with Brown, were he to say "Yes, Jones does eat like a pig," we should think him guilty of insincerity. On the other hand, if Robinson agrees with Brown that Jones habitually wipes his mouth with his sleeve, spits on the tablecloth, and tosses chicken bones on the floor, and if he agrees that Brown has not misapplied the expression "eats like a pig," we may then say that there is a sense in which Robinson does accept and agree with what Brown said, in that he holds the beliefs required for acceptance and agreement.

---

[14] It may here be objected that the fact that Robinson does not share Brown's attitude is no reason to deny that he accepts what Brown said about Jones, on the ground that Brown's attitude toward Jones is no part of what he said, but is rather expressed by what he said. But, the application of the expression "what was said" is not so limited as the objection supposes. Often, what is expressed or implied by an utterance can be counted part of what the speaker said. If you ask me "Is Jones a good philosopher?" and I answer "He is an exceedingly nice person," I imply that I think him a bad philosopher. It would be perfectly correct for you to report our conversation by saying "Part of what Smith said about Jones is that he is a bad philosopher;" and it would be disingenuous of me to protest your saying this on the ground that I had not uttered the words "Jones is a bad philosopher."

The general point here is this. Believing is not the only way in which persons may accept what is expressed by utterances of certain sentences, nor is divergence in belief the only kind of disagreement. Whether we are to say of a particular case that it is a case of acceptance or agreement depends upon whether the criterion for acceptance or agreement we use is that of belief or of assent; and this in turn depends upon the interests which we bring to the case. If, for example, we are merely interested in what Brown and Robinson might honestly answer in a True-False test, then we may say that Robinson believes and accepts that Jones eats like a pig. This is because he might honestly put down "True" after the sentence "Jones eats like a pig." And, given only this interest, we may say that Brown and Robinson agree. But if, as is much more likely, we are interested in how Brown and Robinson will treat and regard Jones, we should say that Robinson does not accept what Brown said and that they do not agree, for it is their attitudes and not merely their beliefs which are our present concern. Given this latter interest, it would be pointless to state Robinson's belief that Jones eats like a pig. Indeed, stating his belief would carry a *suggestio falsi*, for it would imply that Robinson has the attitudes characteristically expressed by "Jones eats like a pig."

The externalist is now in a position to explain why internalism is intuitively plausible. Using the distinction between belief and assent, he may argue: "Sentences such as 'Segregation is morally wrong,' which are typically used to express moral judgments, are also typically used to express particular attitudes. If a person then lacks the attitudes characteristically expressed by utterances of such sentences, he fails to assent to what their utterances characteristically express. Now, the plausibility of internalism lies in the fact that, given our normal moral interests, our criterion of moral acceptance and agreement is assent. In determining a man's moral convictions we are primarily interested in how he will treat himself and others; and so we are not interested merely in what he might profess if asked about moral matters, but we are also interested in his attitudes. If we try to persuade someone that segregation is morally wrong, and if he finally admits that it is but does not adopt an unfavorable attitude toward it, we should no doubt feel that he still does not agree with us about segregation and that he did not really accept what we said about it. It is natural to generalize from such a case and to conclude that moral acceptance and agreement always requires the adoption of a favorable or an unfavorable attitude. But, although reached by a natural route, the conclusion is mistaken. It is true that, if a person is indifferent toward segregation, we should not usually say that he accepts that segregation is morally wrong, even though he may sincerely assert the sentence 'Segregation is morally wrong' and fully understand its meaning. But, that we would not ordinarily say this in no way shows that he does not believe that segregation is morally wrong. By way of analogy, we should not ordinarily say that Robinson believes that Jones eats like a pig if he approves of Jones's table manners; but Robinson may nonetheless believe this. And, were someone to hold that someone must disapprove of Jones's table manners to believe that Jones eats like a pig, his claim would have the same initial plausibility as the claim that the man who is indifferent toward segregation cannot believe that it is wrong. Still, neither contention is true."

It appears, then, that the externalist can explain, consistently with his view, why internalism is such an intuitively plausible position. This, together with the fact that internalists have as yet failed to provide decisive argument for their position, leads me to the conclusion that there is as yet no better reason to favor internalism over externalism. The dispute between internalism and externalism is as yet a stand-off.

IV

I shall now try to show, as briefly as I can, that the absence of compelling argument in favor of internalism has serious consequences for noncognitive metaethical theories. I believe it removes important support from two propositions common to all such theories: first, that such words as "good," "bad," "wrong," etc., have a different kind of meaning than do such words as "green" or "rectangular," and secondly, that moral judgments are not a species of factual judgments.

As for the first proposition, with sufficient space I believe I could show that many noncognitivists have used internalism as a premiss in their arguments that purport to show that "good," *et al.*, have some special kind of meaning. And so, if internalism lacks conclusive support, then any such argument is equally inconclusive. But, since not all noncognitivists have based their analyses of value words on internalism, I shall not press this line of argument. There is, however, another way in which internalism has supported noncognitive analyses of value words. Philosophers have very often, I think, been moved to formulate such analyses because they were convinced that internalism is true and they thought their analyses should reflect this. For example, in Stevenson's early paper, "The Emotive Meaning of Ethical Terms," he sets out three requirements for a satisfactory analysis of "the typical sense" of "good": the second of which is that:

> . . . "goodness" must have, so to speak, a magnetism. A person who recognizes *X* to be good must *ipso facto* acquire a stronger tendency to act in its favor than he otherwise would have had.[15]

Stevenson has simply assumed, without the semblance of an argument, that an adequate analysis of "good" will presuppose or entail the truth of internalism.

The same methodological tendency is operative in R. M. Hare's theory, as this passage from *The Language of Morals* suggests:

> For all of the words discussed in Parts II and III [i.e., "good," "bad," "right," "wrong," "ought,"] have it as their distinctive function either to commend or in some other way to guide choices; and it is this essential feature which defies any analysis in purely factual terms. But to guide choices or actions, a moral judgment has to be such that if a person assents to it, he must assent to some imperative derivable from it.[16]

A person who assents to an imperative, according to Hare, has at least some inclination to act in accordance with it.[17] It is clear, then, that Hare thinks that value words cannot be defined "factually" because he thinks that accepting a moral judgment requires having some inclination to act in accordance with it. And this, in effect, is to hold that value words must be defined noncognitively because internalism is true.

The bearing of my conclusions upon this methodological tendency is plain. If the dispute between internalism and externalism is a stand-off, then until the issue is resolved, we should favor analyses of value words that are consistent with both positions; and we certainly ought not require that a satisfactory analysis of value words presuppose or entail that one of them is true. My conclusions about internalism do nothing to refute noncognitive analyses of value words, but they may perhaps help to remove what has been an important incentive to formulate such analyses.

As for the contention that moral judgments are not "factual," noncognitivists have often appealed to internalism when trying to show this. Indeed, they are sometimes content to assert internalism, and then claim this to mark off moral judgments from factual ones. In the "Retrospective Comments" in his collection of essays, *Facts and Values*, Stevenson proceeds in much this way. He writes:

> There is no reason to suppose, however, that an evaluative sentence is limited to these cognitive functions [i.e., expressing beliefs], and that (as naturalism would have it) it is accordingly indistinguishable from a certain kind of factual sentence. We would not classify it as "evaluative," in my opinion, unless it tended also to express the speaker's attitudes.[18]

By "tending to express an attitude" Stevenson means that, when uttered, an evaluative sentence will express an attitude unless the speaker is trying to deceive, either by lying or by being hypocritical.[19] Since evaluative judgments (of which moral judgments are a species) are presumably expressed sincerely only by sincere utterances of evaluative sentences, Stevenson in effect here argues that moral judgments are not factual because internalism is true. But, until internalism is given conclusive support, this way of distinguishing moral judgments from factual ones cannot be counted adequate.

There are two other arguments which purport to show moral judgments to be nonfactual that rest upon internalism. Both are familiar, and neither requires prolonged discussion. The first is that moral judgments cannot be factual because we cannot be forced to accept them by facts, since to accept a moral judgment is always to adopt a particular attitude.[20] Quite obviously, until it is shown that accepting a moral judgment does require adopting attitudes, this argument never gets off the ground. Secondly, it has been argued here that were moral judgments factual there would be a gap between moral thought and action. Just as we need not be spurred to action by the recognition that something is green, we could then recognize that certain acts are right or obligatory and yet want a sufficient incentive to perform them. We could then say, "I agree that, e.g., keeping my promise is obligatory, but why ought I do it?" And, this latter question is held to be in some way odd or absurd.[21] However, the force of the argument is entirely removed if we take seriously the possibility that internalism is false. If someone may accept a moral judgment without having an inclination to act in accordance with it, then plainly there is at least a logical gap between moral thought and action: and the question "Why ought I fulfill my obligations?" cannot lightly be dismissed by calling it absurd.

---

[15] *Facts and Values*, *op. cit.*, p. 13.

[16] *The Language of Morals*, *op. cit.*, p. 171.

[17] See footnote 5 above.

[18] *Facts and Values*, *op. cit.*, p. 209.

[19] *Ibid.*, p. 206f.

To sum up, noncognitivists have relied heavily upon internalism. Therefore, if I am right in contending that there is no conclusive argument in favor of internalism, then to the extent that they rest upon internalism noncognitivist metaethical theories lack conclusive support.

Smith College

Received June 19, 1970

---

[20] See: Stevenson, *Ethics and Language, op. cit.*, p. 30f.; Hare, *Freedom and Reason* (Oxford, 1963), p. 188f.

[21] See: P. F. Nowell-Smith, *Ethics, op. cit.*, p. 42.

American Philosophical Quarterly
Volume 9, Number 1, January 1972

X. THE PARADOX OF LOGICAL PSYCHOLOGISM: HUSSERL'S WAY OUT

DALLAS WILLARD

I

LOGICAL Psychologism is the view that the non-normative statements made by logicians engaged in their business both are about, and draw their evidence from the examination of, the particular conceivings, assertings, and inferrings of particular persons—a range of facts commonly thought to belong ultimately to the science of psychology alone. This view enjoyed wide acceptation during the last half of the 19th century, being advocated by such men as John Stuart Mill in England, and Sigwart and Erdmann in Germany. More recently, and due as much to misunderstandings and an unfavorable press as to the essential incredibilities it contains, it has suffered an eclipse which is almost total. A part of the purpose of this paper is to show how this eclipse hinders understanding of what Logic is about.

What is here called the "paradox" of Logical Psychologism arises when one sees that, while (i) the class of statements mentioned—we might call them "logicians' truths"—are indeed, in some very important sense, about and applicable to such particular events in personal careers as referred to above, (ii) they nonetheless, as Husserl, Frege, and others have shown, do *not* draw their evidence from the examination of such events. Now this seems paradoxical. For *how* can claims about a certain sort of thing fail to draw their evidence from the examination of things of that sort? There seem to be good reasons for believing both (i) and (ii). And yet the truth of (i) or (ii) seems each to exclude the truth of the other. Thus we have a paradox or antinomy in the classical sense. This paper also proposes to consider a way of resolving the paradox without calling either (i) or (ii) into question.

II

Logicians today are likely to feel uneasy about (i). In this respect we seem to be the inverse of most 19th century philosophers, who easily accepted (i), but had difficulty with (ii). We today have heard or read many snappy announcements that the laws of logic are not expressions of how we actually *do* think, but of how we *ought* to think. And then—with these ringing in our ears—we fail to notice that logicians in fact fill up large volumes without once saying that anyone ought to do anything at all:—Just try counting the "oughts" in Quine's *Methods of Logic*, or Mates's *Elementary Logic*. We fail to see that "valid," "invalid," "tautologous," "consistent," "inconsistent," "derivable," and the other terms which form the core of the logician's vocabulary are not used by him at this trade to commend or condemn, to praise or to blame, to exhort or to direct, at all. That an argument, or argument form, is valid or invalid is a point of mere information or misinformation, to which some positive or negative value may or may not be attached. Value considerations have nothing at all essentially to do with whether or not an argument is valid or invalid, or with whether or not an expression is L-determinate.

It is, of course, true that terms from, or closely related to, the core vocabulary of the logician are often given a normative use—are often used to commend or condemn—in the contexts of actual thinking and talking. An example would be the common claim that a certain person at a certain point in an essay or lecture advanced an invalid (or valid) argument, or on a certain occasion made an invalid (or valid) inference. But it is still true here that the normative use is accidental only. A particular instance of the Barbara syllogism, for example, is valid—i.e., no argument of its sort has or can have true premisses and false conclusion—regardless of what value may or may not be placed upon it; and, of course, also regardless of how anyone's beliefs may be arranged around the premisses and conclusion. *It* is valid, and is known to be valid. That is, the logician's claims (about validity in this case) apply to *it*, the *particular* event of thinking or speaking in the life of a *particular* person. And persons who are logically informed and trained can *know* that *it*, and not merely Barbara syllogisms in general, is valid.[1]

Without this kind of application there would simply be no use at all for what the logician has to teach. If he denies the possibility of such application, he talks himself out of a job. But there is no real danger of that. No logician—even the most hide-bound of anti-Psychologists or Formalists—will assure those whom he would teach that what he has to say is irrelevant to the thinking and speaking which they, or at least *some* persons, may happen to do or meet with. He may have no ready account of *how* it is relevant to actual thinking and speaking, and may—as many teachers and texts do today—either simply avoid the subject, or content himself with a few remarks about its obscurity (or utter impossibility). But that the remarks he makes at his trade are assumed by him to apply to the actual thought and discourse of himself and others is shown by how he *uses* the logical tools which he has sharpened. He *does* analyze particular arguments and inferences, and he does it *with* the principles and techniques which he has developed. Further, he *expects* others—especially if he has taught them—to do the same. No doubt most students of logic fail to know *why* or *how* what their instructor or text has told them applies to actual thinking and talking; and that is probably why only a miniscule percentage of them do in fact ever make *any* use whatsoever of what they learned from their study of logic.[2] But instructors and texts in logic do yet convey to them the assumption that what they are being taught applies to and informs on actual thinking and talking. To this assumption Logical Psychologism did full and explicit justice; but, with very few exceptions, such as John Dewey, logicians of this century have not even faced up to it.

III

But while a logician working today would, perhaps with some uneasiness, agree with Psychologism so far as to own up to (i), he certainly would also reject Psychologism by asserting, with (ii), that the *evidence* which he has for the truths he teaches does not arise from an examination of particular episodes of thinking and speaking. He would assert this simply because he *has* evidence for his truths, but he did not get it in that way prescribed by Psychologism, and because he feels that the *sort* of evidence he has cannot be empirically derived. If Psychologism were correct, then when a logician is speaking, in general and non-normative terms, about propositions, statements, proofs, arguments or inferences, his claims could have only that degree of probability provided by actual *observation* of instances from the classes of psychical or linguistic facts to which he refers. But he very well knows that his claims are not probability statements at all, and that they are never tested by observation of instances.

For example, Quine says, at the opening of "Part One" of his *Methods of Logic*:

> The peculiarity of *statements* which sets them apart from other linguistic forms is that they admit of truth and falsity, and may hence be significantly affirmed and denied. To deny a statement is to affirm another statement, known as the *negation* or contradictory of the first.

If, now, we were to approach this (quite non-normative) passage with a consistently psychologistic attitude, the following sorts of questions would have to be asked: How are these claims which Quine makes known to be true? How did he "discover" them? What are the data upon which these claims rest? When were the relevant observations made? What kind of person was making the statements, the denials, the affirmations, etc., observed? And was the observer qualified or trained in the making of such observations? Did he correctly record his observations? Is a margin of error to be allowed for in such observations? If not, why not? If so, what is it? And was it correctly handled in the observations here in question? What were the conditions under which the observations relevant to Quine's statements were made? Were the conditions sufficiently varied to insure that the correlation to be drawn is indeed just between the classes mentioned (e.g., statements, truths, denials, etc.) and not some others?—*If* non-normative logicians' truths belong to psychology—and perhaps, more specifically, to the psychology of linguistic behavior—and *if* psychology is a science in which evidence is ultimately drawn from observations of fact, *then* these questions *must* be asked. Otherwise there simply is no evidence for the logicians' truths in question. And they must, of course, be asked with reference to *all* of the non-normative truths of logic, not only those stated by Quine above: truths such as, "No proposition (or sentence or statement or utterance) is both true and false," "Arguments with inconsistent premisses are valid, but never show their conclusion to be true," "The universal quantifier validly distributes through a conjunction, but not through a disjunction," and so on. *All* of the evidence for the truth of these and like claims, as well as for their falsity, must ultimately derive from observation of instances of the appropriate classes of experiences, linguistic and otherwise, if Psychologism is true. But, of course, it simply does not do so. Hence, Psychologism is false. It is wrong about where and how the logician obtains evidence for the truths he teaches.

---

[1] Quine, for one, explicitly says that what admits of truth and falsity—and, by implication, of logical relations—are "individual events of statement utterance." (*Methods of Logic*, rev. ed. (New York, 1959), p. xi.) Unfortunately, this statement forms no homogeneous part of his logical theories, but rather is forced upon him by his theory *about* logic, and his general epistemological views. One who knew only what Quine has said in developing logical theories would not have the least inkling that the theories were theories about utterance events.

[2] As a matter of simple fact, few courses more totally—though perhaps many equally—waste the student's time, so far as his own self-betterment is concerned, than university courses in logic. The irony is that the greater waste is usually made by the very courses which are technically more rigorous. A nice, sloppy course mainly dealing with so-called "informal fallacies" often gives the student a critical handhold (usually involving little *insight*) on some things he subsequently thinks or hears.

IV

All of which leaves us with the puzzle of how (i) and (ii) can be shown consistent with each other. It is interesting that many philosophers seem not to have been aware of this puzzle, much less worked out any solution for it. Frege, for example, distinguishes very clearly between what he calls an "idea" (a temporalized segment of the experience of some particular person) and what he calls a "thought."[3] As is well known, he takes certain complex 'thoughts" to be the bearers of truth, and assigns "to logic the task of discovering the laws of truth, not the laws of asserting or thinking."[4] But as to an elucidation of *how* these "thoughts" and "ideas" are related to one another, especially in such a way that the laws of the former allow us to ascertain (as we constantly do) the logical character of the latter, e.g., valid, inconsistent, etc.—on this Frege has nothing helpful to say. He agrees that something in the individual consciousness must be "aimed at" the thought for us to think it or have it.[5] But he has no analysis of this metaphor of "aiming." He gives the name "apprehension" (Fassen) to the relation of mind or "idea" to the thought, but of course has to acknowledge that this too is a metaphor.[6] The puzzle of how logicians' truths apply to and inform us about certain aspects of actual thinking and asserting is left unresolved by Frege, if, indeed, he had any clear perception of it at all. And in this respect he is typical of many of the best logicians and philosophers of recent decades.

Husserl, on the other hand, was driven to work out a solution to this puzzle by his attempts to understand, from what was at first a strongly psychologistic point of view, the anti-Psychologistic view of the *proposition*, as presented by Bolzano.

The pattern for the standard, anti-Psychologistic view of propositions (and concepts as well) was set by Bolzano's "*Satz an sich*," established by Frege's "*Gedanke*," and naturalized in Anglo-Saxon countries in the "proposition" of Russell, Moore, W. E. Johnson, L. S. Stebbing, R. M. Eaton, and many other writers. In this pattern (and it is necessary here to think of a *pattern*, which allows some variation in the way it is specified by each writer) a proposition has the following main features:

1. The proposition is not located in space or in time, as, for example, sentences, or events involving sentences, are. This is in part what is meant by saying that they do not have *actual* or *real* existence, but only *ideal being*.
2. The proposition is not identical with a sentence, but the meaning or sense of an (indicative) sentence is a proposition. Propositions may, thus, be "expressed" by sentences, though it is not *essential* to them that they be so expressed.
3. The proposition is not something which can be sensuously perceived—even as a faint "glow" upon meaningful sentences or "living" words—though it *is*, of course, *somehow* known or, to use Frege's term, "apprehended."
4. The *same* proposition can be somehow

---

[3] "The Thought: A Logical Inquiry," in P. F. Strawson, ed., *Philosophical Logic* (Oxford, 1967), pp. 26–29; and "Der Gedanke," in Gottlob Frege, *Logische Untersuchungen*, Günther Patzig, ed. (Göttingen, 1966), pp. 40–43.

[4] Strawson, *op. cit.*, p. 19; and Frege, *op. cit.*, p. 31.

[5] Strawson, *op. cit.*, p. 35; and Frege, *op. cit.*, p. 50.

[6] Strawson, *op. cit.*, p. 35n; and Frege, *op. cit.*, p. 49n.

grasped by the minds of *many* persons. In this capacity the proposition has been invoked as the intersubjective basis of the phenomena of communication and of the objectivity of scientific knowledge.

5. But it *need not* be grasped by the mind of any person at all. Its *esse* is not *percipi*. As Frege says, "In thinking we do not produce thoughts but we apprehend them . . . The work of science does not consist of creation but of the discovery of true thoughts."[7]
6. When the proposition is related to a mind, its relation is, or principally is, that of an *object* of thought or of the so-called "propositional attitudes," such as belief or doubt. It is "before" the mind. This is sometimes obscurely expressed by saying that it is the "content" of a belief or judgment.
7. Description of a proposition does not *essentially* involve a reference to any particular mind or act of thought with which it may be involved. Its truth value or what it refers to is never affected by such involvement, unless it happens to be a proposition which *refers* to such involvement.
8. But its description does essentially involve mention of its references *to*, or intendings or meanings *of*, certain things, plus description of how these references are related to one another. A proposition essentially consists of a more or less complex set of references, intentions, or meanings, plus the mode of their combination.
9. Finally, the proposition is what is underivitively true or false, while opinions or sentences or statements are true or false only because they have a certain relationship with a proposition.

It would not be worthwhile here to document each point of this pattern in all of the writers mentioned above. The whole pattern appears in subsections 19, 25–26, 34, and 48–50 of vol. I of Bernard Bolzano's *Wissenschaftslehre*. In Frege's paper on "The Thought," already referred to, the pattern once again appears whole. It is almost whole in Moore's *Some Main Problems of Philosophy*, chap. 3; in chap. 1 of W. E. Johnson's *Logic*, Pt. I; on pages 6–24 of R. M. Eaton's *General Logic*; and on pages 17–20 of Max Black's *Critical Thinking* (2nd edition). Bertrand Russell's "On Propositions: What They Are and How They Mean," in his *Logic and Knowledge*, as well as chap. 12 of his *Analysis of Mind*, presents important phases of the pattern, as well as some modifications of it. So does C. I. Lewis in chap. 3 of *An Analysis of Knowledge and Valuation*. Husserl's *adaptation* of the pattern is nowhere presented more clearly than in chaps. 7 and 8 of Vol. I, and in subsections 29–35 of the first "Investigation" in Vol. II of his *Logische Untersuchungen*.

VI

And it obviously is with an *adaptation* or *modification* of the anti-Psychologistic pattern that we must be concerned, if its concept of the proposition is to be used in a resolution of the paradox at hand. It is clear that, if logicians' claims are fundamentally about the "propositions" described, then evidence for those claims could not be derived from a mere inductive survey of particular thinkings and speakings, to which, in the pattern, propositions have only a quite extrinsic and contingent connection. Thus, the pattern supports our (ii), and Husserl strongly concurs with it in this respect. But it is equally clear that the pattern, as it stands, gives us no help in reconciling (ii) with (i). It gives not the least suggestion of how logicians, having the source and sort of evidence they do have for their claims, do nonetheless succeed in saying something about the actual thinking and speaking of particular human beings. Where in the anti-Psychologistic pattern does the problem lie?

According to Bolzano—and point six of the anti-Psychologistic pattern—the relation between the mind and the propositions which it "thinks" or "grasps" is that of *intentionality*. That is, the proposition is the mind's *object*; or, to use one of Frege's terms, it "stands over against" (*gegenübersteht*)[8] the mind thinking it. L. S. Stebbing expresses this view of the proposition's relation to mind in words frequent in logical writings of her time, by saying that "a proposition is anything that is believed, disbelieved, doubted or supposed."[9] Of course the objections to this position on propositions are quite well known. They are stated, for example, on pages 35–36, of Vol. I of McTaggart's *The Nature of Existence*, and more fully on pages

---

[7] Strawson, *op. cit.*, p. 35; Frege, *op. cit.*, p. 50.

[8] In "Die Verneinung: Eine Logische Untersuchung," in Gottlob Frege, *op. cit.*, p. 58.

[9] *A Modern Introduction to Logic* (London, 1961), p. 33. Cf. pp. 4–7 of W. E. Johnson, *Logic*, Pt. One (New York, 1964).

105–111 of Gilbert Ryle's paper, "Are There Propositions?"[10] Perhaps the main points in these objections may be summed up by saying that the alleged "objective" propositions are a queer, or at least superfluous, sort of entity, and that they *are not* always, if they are ever, the objects of our beliefs, doubts, etc.

In fact, these are precisely Husserl's objections to objective propositions. He thought of Bolzano's "propositions in themselves" as "mythical entities, suspended between being and non-being."[11] And he held that when "we make an assertion, we judge about the fact concerned, and . . . not about the judgment in the logical sense,"[12] i.e., about the proposition. However, what is principally of interest here is not Husserl's objections to objective propositions, but rather his way of avoiding those objections, while still retaining the essence of anti-Psychologism in *his* view of the proposition and, simultaneously, resolving the paradox of Logical Psychologism.

VII

In the review just quoted from, Husserl indicates that the key for his revision of Bolzano's theory of propositions came from his study of yet another revision: Lotze's revision of Plato's theory of forms.[13] Plato's forms had, like Bolzano's propositions, been persistently called both queer and useless; and the queerness and uselessness alleged seemed mainly to follow upon their being *located in* something at some sort of "distance" from their instances—in some void or Divine Mind or *Τόπος οὐράνιος*. Lotze saw that this odd kind of "localization" was not required by the various arguments for universals, and that it indeed made no sense anyway. He held that to say that blueness or triangularity, for example, "exists" or "is" is not to say that a certain color or plane figure is located somewhere, nor is it to imply this.[14] It is only to say that certain discriminable elements of things have certain (always non-spatial and non-temporal) determinations of their own:—e.g., blueness is a color which is darker than yellowness (though no blue thing is), and triangularity is a plane figure involving fewer straight lines than squareness (though no triangular object is). One—really not quite correct—way which Lotze (and later Husserl) had of saying this was to say that for universals to exists means only that certain propositions about them are true.[15] To mark this point of connection between their existence and truth, Lotze said that universals do not have *real* existence, or *Wirklichkeit*, but only *ideal* being, which he also called *Geltung* or *validity*.[16]

Now Husserl does not detail for us precisely *how* this revision of Plato's theory of forms led to his revision of Bolzano's theory of the proposition "in itself." One can see, however, an attractive route which he may have traveled:—If universals can avoid the charges of queerness and superfluity by the assignment of ideal being or *Geltung*, then possibly propositions can escape by being regarded as universals. Such an identification of propositions with universals (of a certain sort) must have been rendered more attractive by observing that it conflicted with the anti-Psychologistic pattern for the proposition only on point six, and that the other points either do not conflict with it or actually seem to strengthen it, as does point four.

In any case, Husserl came to regard the propositions of which logicians speak, and for the truth of which, according both to him and Frege, they discover laws, as complex, *referential qualities*, which may or may not be instanced in minds, but which —following Brentano—are never instanced by any physical thing. He said that upon reading Lotze he

> . . . saw that under "proposition in itself" is to be understood, . . . what is designated in ordinary discourse—which always objectifies the ideal—as the "sense" of a statement. It is that which is explained as one and the same where, for example, different persons are said to have asserted the same thing. . . .

---

[10] *Proceedings of the Aristotelian Society*, vol. 30 (1929–1930), pp. 105–111.

[11] From page 290 of Husserl's review of *Der Streit der Psychologisten und Formalisten in der modernen Logik*, by Melchior Palágyi (Leipzig, 1902). This review appeared in vol. 31 (1903) of *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane*, pp. 287–294.

[12] *Logische Untersuchungen*, 5th ed., vol. 2, Pt. One, p. 103. An English translation of this review is to appear in *The Personalist*, for 1972.

[13] Hermann Lotze, *Logic*, edited and translated by Bernard Bosanquet (Oxford, 1884), Bk. III, chap. ii, pp. 433–449.

[14] *Ibid*, p. 443.

[15] *Ibid*, pp. 439–446. And for Husserl see *Logische Untersuchungen*, Vol. II, Pt. I, p. 101.

[16] Of course, this is no mere *in rebus* or *post rem* doctrine of universals. To deny that universals exist in some place apart from their instances is not at all to hold that they exist only (or at all) in their instances or in minds which have beheld their instances. Nor is it to say that they in any way depend, for their being or their being known, upon their instances.

> And it further became clear to me that this identical sense could be nothing other than the universal, the species, which belongs to a certain *moment* or phase present in all actual assertions (Aussagen) with the same sense, and which makes possible the identification [of sameness] just mentioned, even where the descriptive content of the individual experiences (Erlebnisse) of asserting varies considerably in other respects. The proposition (Satz) thus relates to those acts of judgment (Urteilsakte) to which it belongs as their identical intention (Meinung) in the same way, for example, as the species *redness* relates to individuals of "the same" red color. Now with this view of things as a basis, Bolzano's theory, that propositions are objects which, nonetheless, have no "existence" (Existenz), comes to have the following quite intelligible signification:—They have the "ideal" being (Sein) or validity (Gelten) of objects which are universals (allgemeiner Gegenstände); and, thus, that being which is established, for example, in the "existence proofs" of Mathematics. But they do not have the real being (reale Sein) of *things*, or of dependent, thing-like *moments*—of temporal particulars in general. Bolzano himself did not give the faintest intimation that these phenomenological relationships between signification, signification moment, and full act of signifying had been noticed by him.[17]

And elsewhere Husserl says:

> . . . In the actual experience of signification there corresponds to the unitary signification[18] an individual aspect, as singular instance of the signification species: in the same way as in the red object the red-moment corresponds to the specific difference, red. If we actualize the act of signifying and, as it were, live in it, then we naturally mind the act's object and not its signification. When, for example, we make an assertion, we judge about the fact concerned, and not about the signification of the indicative sentence involved, not about the judgment in the logical sense. This latter becomes an object for us only in a reflective act of thought, in which we do not merely look back upon the *real*-ized assertion, but rather carry out the required abstraction (or, better said, ideation). This logical reflection is not . . . an act which occurs under certain artificially induced conditions. Rather, it is a normal constituent of logical thinking.[19]

We may, therefore, sum up Husserl's revision of Bolzano's view of the proposition as follows: For Bolzano (and most anti-Psychologists) the relation of mind to proposition is *intentionality*, while for Husserl it is *instantiation* or *exemplification*.[20] The obviously increased intimacy of the connection as viewed by Husserl is the essence of his resolution of the paradox of Logical Psychologism.

VIII

It seemed paradoxical that logicians could succeed in saying something true of particular events of thinking and speaking, since they do not develop their theories by inductive or empirical analysis of such events. Given Husserl's revision of the concept of the proposition, this fact may be explained as follows: Propositions, on his view, are not particular acts of thought, but complex, referential characters or qualities *of* such acts. These referential qualities also have determinations, among which are truth and falsity and conditions of truth and falsity. But these determinations do not merely have some quite extrinsic and contingent connection with the concrete acts of thinking and speaking which instance their immediate subjects, i.e., propositions. Somewhat as, since red is a color, a red thing, though no color, is necessarily a colored thing, so, we might say, an act of thought or speech which instances a true proposition, though itself no truth, is nonetheless a true judgment or utterance. And the various logical relations which hold between the truth conditions of propositions can similarly be given a natural extension to apply to the acts of belief or inference which involve instantiation of the propositions in question. The proposition, or set of propositions, introduces its determinations (properties and relations) into the individual acts which instance it.[21]

Because of this connection between universals and their instances, truths about universals, including, of course, truths about concepts and propositions,[22] entail corresponding truths about the corresponding individual things or events.[23] To take examples from logic, the proposition that the proposition that Nixon is a Republican is true, has its transform in the proposition that all judgments or statements to the effect that Nixon is a Republican are true. Or again, the proposition that the premisses of a certain syllogism of type Barbara imply its conclusion has its transform in the proposition that all actual inferences from those premisses to that conclusion are valid.

---

[17] From p. 290 of the review of Palágyi.

[18] A "proposition" was one type of "signification" (Bedeutung) for Husserl, as it was one type of "sense" for Frege.

[19] *Logische Untersuchungen*, Vol. II, p. I, p. 103.

[20] The similarity between the views of Husserl and those of Gustav Bergmann on propositions should now be obvious. See Bergmann's *Logic and Reality* (Madison, 1964), p. 34 and elsewhere. The American psychologist R. S. Woodworth also advocated similar views early in this century. On p. 706 of his "Imageless Thought" (in *The Journal of Philosophy* for 1906) he says: "There is a specific and unanalyzable conscious quale for every individual and general notion, for every judgment and supposition. These qualities recur in the same sense as red and green recur."

[21] *Logische Untersuchungen*, 5th edition, Vol. I. See the sentence spanning pp. 142–143 and the sentence ending the paragraph on p. 151.

What the logician's truths are *primarily* about are, therefore, not mental or linguistic acts, but characters of acts; and it is from the analysis of the truth conditions of these characters (which, as universals, are always the same, no matter what their instances may be) that he obtains evidence for his claims. But his truths also apply informatively to particular acts, and do so precisely because they are about the characters *of* such acts. And *this* is Husserl's way out of the paradox of Logical Psychologism. It is unfortunate that it has been so universally neglected or misunderstood; for it concerns a central issue in the philosophy of logic, and there is perhaps no other philosopher who has worked out a way of reconciling our (i) and (ii) that is remotely as plausible as Husserl's way.[24]

University of Southern California

Received June 10, 1970

---

[22] On p. 342 of *Logische Untersuchungen*, Vol. II, Husserl says: "As to all ideal units [i.e., universals], so there correspond to significations real possibilities and, perhaps, actualities. To significations *in specie* there correspond acts of signifying; and, indeed, the former are nothing but aspects of the latter, taken as ideal (*als Ideal gefasste*)."

[23] See subsections 7 and 8 of Husserl's *Ideas* (New York, 1962), for elaboration of this claim. A clear exposition of essentially the same position is found in H. W. B. Joseph, *An Introduction to Logic*, 2nd ed. (Oxford, 1916), pp. 282–285.

[24] It must be added that the above account is by no means the whole story on Husserl's view of propositions. In *Ideas*, Vol. 1 (see especially subsection 93) and in *Formal and Transcendental Logic* (The Hague, 1969), one finds an immensely—and, I think, unfortunately—more complex theory of propositions, which returns to an essentially Bolzanian or "neomatic" proposition.

American Philosophical Quarterly
Volume 9, Number 1, January 1972

XI. WHAT'S HAPPENING IN PHILOSOPHY OF LANGUAGE TODAY—A METAPHYSICIAN'S-EYE VIEW

JAY F. ROSENBERG

CONTEMPORARY philosophy of language is dichotomized. There are those who talk about *regularities* and there are those who talk about *rules*. On the regularities side, we have, for example, Ziff, Quine, and sometimes Wittgenstein; on the rules side, Searle, Chomsky, and, again, sometimes Wittgenstein. The dichotomy gives rise to oppositions of the following sort:

> Speaking a language is engaging in a (highly complex) rule-governed form of behavior. To learn and master a language is (*inter alia*) to learn and have mastered these rules. (Searle; *SpA*, p. 12.)[1]

> I am concerned with regularities: I am not concerned with rules. Rules have virtually nothing to do with speaking or understanding a natural language. . . . An appeal to rules in the course of discussing the regularities to be found in a natural language is as irrelevant as an appeal to the laws of Massachusetts while discussing the laws of motion. (Ziff; *SA*, pp. 34–35.)

If you listen with an Hegelian ear, I think you will hear a thesis and an antithesis crying out for synthesis. It's a lot like a typical marriage: They can't live with each other, but they can't live without each other either. Also, as a metaphysician, I hear echoes here. The debate has a ring of familiarity about it. I should like, however, to postpone the echoes for a while. Let me try some marriage counseling first.

Why can't they live with each other? Let's pursue Ziff's case against rules a little further.

> It is possible to misuse words, or to use them correctly or incorrectly; e.g., philosophers who speak of "the rules of language" (or of "moral rules") are I believe misusing the word "rule." But I do not think they are breaking any rules. It is possible to misuse a screwdriver, to use it correctly or incorrectly: are there rules for using screwdrivers? This too would be an odd use of the word.
>
> If there were "rules of language" then presumably such rules would be laid down in the course of teaching the language. This is another confusion: one is not taught one's native language, one learns it. In our culture a child is not taught to speak at school: he can speak before he goes to school. If he cannot, he is not likely to be taught much at school. If he can speak before he goes to school, who taught him to speak? The parent who teaches his child an occasional word may think he is teaching his child to speak. But a normal person during the early part of his life learns a thousand words a year, more than three words a day on the average. Who teaches him these words? (Ziff; *SA*, p. 35.)

Now there is certainly a good bit here that is true. Misuses of language are more like misuses of tools than like violations of the rules of a game. (And now you should see in part why Wittgenstein comes down on both sides of the debate. He uses *both* analogies.) Speaking a language looks more like pursuing a teleological activity than like playing a game. What I am calling teleological activity is activity directed toward getting something done, activity which is means toward some end. Now there can be, and frequently is, teleological activity within a game context. In seriously played chess, for example, each move is directed toward the eventual end of checkmating the opposing king. Yet the activity of playing chess need not be, and frequently is not, *itself* teleological activity. Chess is often played for its own sake, as an *autotelic* activity. The majority of utterances of a language are teleological, directed toward one or another illocutionary or perlocutionary end—informing, warning, persuading, threatening, insulting, requesting, and the like. There are, of course, autotelic linguistic activities as well. Much, though not all, poetry will fall under this heading. What Ziff is calling attention to is the way in which these categories interact with various terms of criticism. There are two sorts of mismoves one can make at chess. One sort impedes the teleological. One may, for example, leave one's queen *en prix*. This is poor chess, but it is still chess. Not so with what we term a violation of a chess rule. If one starts moving rooks diagonally, one has *a fortiori* ceased to play chess. (One may be playing what is called "fairy chess.") It is not just that the teleological has been impeded. The autotelic has been suspended. Presumed rules of language behave more like the strategical precepts of sound teleological chess than the rule constraints of autotelic chess. To deviate significantly from the norm in language is to fail to be understood, to fail to inform, warn, persuade, or what have you. But it is not to cease speaking or even to cease speaking English. For it may be precisely to pursue autotelic linguistic activity. Thus Ziff quotes (*SA*, 36): "My heart in hiding/ stirred for a bird,—the achieve of, the mastery of the thing!" "Achieve," of course, is not a noun; no adequate grammar would so mark it. Yet we have here a bit of English language poetry; indeed, even a bit which we can understand. It is not here to the point—it is apparently *never* to the point—to remark "You can't say *that*; that's against the rules!"

---

[1] For citations of this form see references at end of the paper.

Ziff's second line of criticism goes a bit deeper. The suggestion seems to be that it is just the fact that speaking a language is *not* obeying rules which makes it possible for a child to learn to speak *without* being taught. The point might be amplified as follows: Learning to behave according to certain rules is learning to pursue or eschew certain activities because they are enjoined or proscribed by the rules. This requires that, in some sense, the rules in question be "present to the mind." Now a rule which enjoins or proscribes a practice must somehow secure reference to that practice, and as the practices here in question are linguistic, the rules in question must presumably have a metalinguistic character. In some sense, then, learning to use a language—on this model—will presuppose having learned to use a metalanguage. As Sellars puts it,

> . . . it is as though when asked "How did German words come to be meaningful to Schmidt?" someone were to say, "Well, before learning German he knew English—though not to speak out loud—and his compatriots, by a clever combination of gestures and the production of vocables in the presence of objects, brought him to formulate to himself (in English) and obey such rules as 'red objects are to be called *rot*'." (*SRLG*, pp. 334–335.)

And this sort of account of language learning clearly will not do at all. Rules, apparently, can only be learned as rules by being taught. The point is a general one about normatives. One cannot, it seems, learn what is enjoined or proscribed simply by observing what is done. One can learn manners by observing and emulating; one must be taught morals. Yet children are not taught a language, but rather they simply learn it. Hence learning a language cannot be a matter of learning rules.

So much for the case against rules. What Ziff would put in their place is regularities. But there are obstacles here.

> If "The cat is on the mat" is uttered then generally a philosophicogrammatical discussion is under way. (*SA*, 60.)

That is the regularity. Yet it is clearly irrelevant. Ziff recognizes this. And so he gives us in addition to regularities both *principles* and *projections*. The projection required here is this: That if "The cat is on the mat." is uttered then, in the standard case, some feline is on some mat. The principle is the principle of composition:

> Roughly speaking, it is a principle to the effect that the relevant similarity between distinct semantic correlates of u_i and u_j be a reflection of the relevant similarity between the two utterances. (*SA*, p. 62.)

In a standard case, some feline is on some mat. But what is the sense of "standard" here? One thing is clear, a standard case will not be a case which is *standardly found*, a typical or statistically normal case. Indeed, the standard case may never occur or have occurred. Rather it will be a case against which others are to be measured for deviation, a criterion or norm. It will be the case to which one *ought* to conform to achieve a wide variety of illocutionary or perlocutionary ends. And we cannot have language without such standards. Why not?

When one has acquired a language, one has, *inter alia*, acquired the ability to understand utterances, produced by others, which one has never heard before and to produce utterances, understandable by others, which one has never produced before. While one's own linguistic resources are finite, the potentialities for their deployment are infinite; while one's own linguistic experiences are unique, one's potentialities for speaking and understanding are shared. The great strength of the rules advocates has always been this: that it is utterly unintelligible how diverse speakers of a language can issue and understand utterances which have not occurred before unless we grant that the speakers have in common a set of rules for producing and interpreting utterances. Only in this way can finite resources be *uniformly* projected into infinite combinatorial possibilities. Only if the combinatorial possibilities of language are governed by shared rules *can* language be an instrument for the attaining of illocutionary and perlocutionary ends. If language is an instrument, then, it is an essentially combinatorial instrument. This, apparently, even the most obdurate opponents of rules admit:

> To deny the principle of composition would, I am inclined to suppose, be utterly absurd. Either the new utterance is arbitrarily explicated then and there or the explication is at least partially predetermined by the syntactic structure of the utterance. If every new utterance had then and there to be arbitrarily explicated, the corpus of utterances would not even resemble a language. (Ziff, *SA*, pp. 61–62.)

Yet the only alternative to arbitrary explication would seem to be explication according to some rule. Philosophers of Searle's persuasion conclude, then, that one has acquired a language when and only when one has acquired this shared structure of rules, when and only when one's own linguistic behavior has come to accord with them. So runs the case for rules.

Thus they can't live with each other, but they can't live without each other either. We need something which is *like* a rule of a game in that it has normative force and can guide or channel or structure behavior in such a way as to uniformly project diverse finite resources into the infinite combinatorial possibilities of a natural language, yet which is *unlike* a rule of a game in that it is not constitutive of the activity it regulates but can be freely suspended without suspending *a fortiori* that activity and can be learned without being taught. Or, to put it in a way which gets to the heart of the matter, if language is to be governed by *rules* in any familiar sense of "rule," we need a sense of behavior *governed by* a rule which is stronger than behavior which merely, by happenstance, *conforms to* the rule yet weaker than behavior *in obedience of* the rule. Is there such a thing?

I think that there is. There are various activities which have an arbitrariness about them, which can be done in one way or another indifferently, but which, if it is important that everyone engaged in the activity do it the *same* way, come to be structured into patterns. Take something like driving on the left or the right side of the road, for example. Imagine an absence of specific legislation. *Which* side of the road comes to be the side on which everyone drives is arbitrary; that everyone drive on the *same* side is not arbitrary. It has, if you will, survival value. And so there is room for evolutionary processes to operate. I speak analogically here, but it is a strong analogy. In some cases, it can become the literal truth. Take a case cited by Sellars (*SRLG*, 326–327), the dance of the bees. *How* a bee wiggles is arbitrary; all bees having the same wiggling production and interpretation program is not arbitrary. It has survival value. As such, the mechanism has been caught up in the evolutionary processes. What began as random movements have become parts of a structure of movements, parts of the dance. It is the program for the structure as a whole which has survival value. It is that program which, as the result of the operations of evolutionary selection, has come to be genetically transmitted. And now a clever apiologist comes along and describes the dance in a normative way, lays out a set of rules for the various wigglings and buzzings. Does the bee *obey* these rules? Surely not. Does his behavior then merely conform to them by happenstance? No, not that either. For the rules may be viewed as constitutive of a structure, and individual wigglings and buzzings are properly to be explained by appealing to that structure of which they are projections. So it can be with traffic. Still assuming the absence of specific legislation, human beings can take evolution one step further. When the accident comes, the explicit normative may come with it. "He *ought* to have been driving on the right," we may say. And we may be right. For that's not merely how things *are* done around here; it's how they *are to be done*.

In the acquisition of language by a child, I think, we have an analogue of evolution in miniature. What begin as random phonemes get structured into larger and larger linguistic units according to patterns shared by speakers in the child's immediate environment. The results have survival value. Ultimately some clever linguist may formulate a set of rules for, say, English. Do native English speakers *obey* these rules? There are, as we have seen, good reasons to say that they do not. But do they then merely *conform* to them by happenstance? No, not that either. For the rules may be viewed as constitutive of a structure, and individual phonetic productions *are* properly explained by appealing to that structure of which they are projections.

Where the pattern-governedness is caught up genetically, we have evolution proper. In persons, we do not have—or do not obviously have—genetic involvement of this sort. (On this turns the question of *linguistic universals* in the sense of

Chomsky, ATS. See, e.g., pp. 27–28, 35–37.) Yet we have, in any case, an analogue to evolution and an analogue to rules. And by means of the former process, the latter rule-analogues can be learned without being taught. There are several good words for these rule-analogues. An old one is "custom," in the Humean sense of the term. A more recent one is "convention." One of the things that is happening in philosophy of language today is the emerging recognition that a natural language is governed by conventions in just this sense, and that we may here find one rapprochement between the dichotomous camps of rules and regularities. In this regard, an important book has just appeared. It is by David Lewis, and it is called, interestingly enough, *Convention*.

So much for Hegelian tensions. If that were all that is going on, heaven knows it would be exciting enough. But we're forgetting the echoes. In fact, a drama much larger than this parochial dispute is currently being played out on the stage of philosophy of language. To get a glimpse of it, it may suffice to remind you of an earlier dispute couched in much the same terms. Recall, if you will, the great question of *causality* and the dichotomy of constant conjunction vs. necessary connection. Contrast Humean experiential regularities with physical entailments. Recall, too, that Chomsky has characterized his debate with Quine as a dispute between rationalism and empiricism, and I think you will see that contemporary philosophy of language is very much in classical *medias res*.

Or consider Ch. II of Quine's *Word and Object*, the thesis of the indeterminacy of radical translation. Recall that the term translations projected by a set of analytical hypotheses come in only as a calculating device. Any set of analytical hypotheses which projects native utterances into linguists' utterances with matching stimulus conditions will do, and no such set is to be preferred to any other. Quine insists that the phenomenon is peculiarly linguistic, but, as a metaphysician, I hear an instrumentalist berating a scientific realist: "Talk about electrons is a convenient calculating device for predicting observable experimental outcomes given observable experimental setups," he says. "There is no right or wrong about electrons, no true or false. The only question is whether electron-talk has utility for predicting correlations of observables." Or, finally, take the natural extension of referential indeterminacy to individual idiolects or, indeed, to temporal slices of a single speaker's idiolect, pursued by Quine in his essay "Ontological Relativity." Referential solipsism of the present idiolect reminds me too much of epistemological solipsism of the present moment to let it pass. Or again, and I promise this *is* the last, contrast the problem of other minds with Quine's problem of other languages—and consider the fate of the argument from analogy in both cases.

So we have echoes aplenty. Philosophy of language emerges as merely the latest arena for a perennial philosophical combat. It has been fought in metaphysics over causality, in philosophy of science over theoretical entities, and in philosophy of mind over the mental states of others. The stakes are high, for the issue is nothing less than the adequacy of empiricist epistemology. Now I believe that empiricist epistemology, properly understood, *is* adequate, and I think that the outlines of a defense of it, in all of these various arenas, have been manifesting themselves for some years. To get at the groundworks of a solution, however, it is necessary first to slightly recharacterize the debate. Let me, thus, take a quasi-detour into the philosophy of science.

The traditional account of the methodology of science, what I shall be calling the old orthodoxy, runs roughly as follows: The *goal* of science is the prediction (and control) of natural phenomena. To get into the epistemologically best position for making predictions, we are to use the hypothetico-deductive method. Accounts of the application of the method typically begin with an injunction to gather data. Once we have a bunch of data, we are to form an hypothesis which *fits* the data, either by abstracting or generalizing from the data itself, or by divine inspiration. In any case, once we have an hypothesis, however arrived at, which fits the data in hand, we must proceed to test it for acceptability. To test an hypothesis, one *deduces* predictions. One then observes to determine whether what has been predicted occurs. If the prediction is fulfilled, we are to say that the hypothesis is *confirmed* and provisionally accept it. The more test predictions which are fulfilled, the better confirmed the hypothesis is and the more acceptable it becomes. This is called "induction." If, on the other hand, the prediction is not fulfilled, we are to say that the hypothesis is *falsified* and reject it. We must then search for another hypothesis. There are, of course, many variations on these themes, but this, in outline, is the substance of the old orthodoxy.

It is now, I think, becoming generally recognized that this account is not only moth-eaten, but myth-eaten as well. A new view of scientific methodology is emerging from the work of such men as Duhem, Kuhn, Hansen, Feyerabend, and Sellars. It is a view based upon science as it is and has been practiced, rather than upon philosophers' theories of how science ought to be practiced. The key to this new empiricism is the recognition of a third kind of good reasons, a third mode of empirical inference. The old orthodoxy offers us two—deduction (derivation of predictions) and induction (confirmation of hypotheses). But there is a third. Harman (*IBE*, *EI*) has conveniently called it "inference to the best explanation" and has theorized that it is the key to understanding enumerative induction as well. I think that he is right in this, but for our purposes it will be enough to see its autonomous operation in postulational theory formation. Inference to the best explanation may be summarized by the following maxim: The test of an account of empirical phenomena is the extent to which it *explains* what we already know about those phenomena. Thus we must begin our theorizing, not in ignorance, but with some knowledge of how natural phenomena behave, some set of prior empirical laws. These may be the substance of some predecessor theories, mere common sense generalizations, or carefully established by pursuing something like the hypothetico-deductive method as characterized by the old orthodoxy. There is a perfectly good name for the last process as well, by the way. It is called "natural history." Thus, for example, we may have already in hand the Boyle-Charles Gas Law or Kepler's Laws of planetary motion and Galileo's Law of falling bodies. Now, although such laws are adequate for making highly reliable predictions concerning observable phenomena, the aim of science is not *merely* thus to be able to predict the course of nature. Science rather addresses itself to the further central goal of explaining and understanding why natural phenomena conform to these regularities, why, as it were, they obey these laws. To this end, we typically construct a story, a theory, about unobservables—molecules, for example, or gravitational forces. The test of this theory will be, according to the maxim elucidated above, the extent to which it explains the laws we already have. But this process, the process of explanation, is generally a much more complicated matter than the old orthodoxy—say, Hempel—would have it be. Typically, explanation will involve both idealizations and limit concepts. On the old orthodoxy, a theory explains a law if and only if it entails that law. But, for example, the Boyle-Charles Law is literally false on the kinetic theory, for the molecules posited by that theory do not have zero diameter. And, again, according to Newtonian mechanics, the planetary orbits are *not* ellipses, for a planet is subject to perturbational gravitational forces from other planets, satellites, and interplanetary gas and dust as well as to the central gravitational force of the sun. On the hypothetico-deductive account that would be the end of the matter, for in one sense our newly minted theory does not even *fit* our pre-theoretical information. But what happens in the real world of practicing scientists is something quite different. Each theory may be used to calculate the *magnitude* of the deviation. And each theory has as a consequence the result that, while its predecessor laws *are* literally false, they are true in the limit and it is a limit which we closely approximate. Molecules have *near* zero diameters and the perturbational forces acting upon the planets are *near* vanishingly small. Thus, in general, theories do *not* explain prior empirical laws by entailing them. In one sense, the new theory does not explain the old laws at all, for if the new theory is right, the prior laws are literally false. But what the new theory *does* do is explain the *success* of the prior laws. In Sellars' nice phrase, theories "explain empirical laws by explaining why observable things obey, to the extent that they do, these empirical laws" (*LT*, 121). If a theory accomplishes *this* job, then we have a good reason for accepting it, and a good reason, *inter alia*, for believing that the unobservables postulated by the theory are true of the world. And if the theory accomplishes this job, predictive adequacy may be very much not to the point. When Uranus is discovered to deviate from its predicted position, Newtonian mechanics is not imperiled. Rather Newtonian mechanics is retained and, in fact, *appealed to* in an effort to explain the deviation. In the case of Uranus, the appeal works. Neptune shows up on schedule. In the case of Mercury, by contrast, it doesn't work. The 40′ of arc deviation then hangs around for years as an unexplained anomaly. But Newton's theory is not thereby shaken, for *any* theory fits *any* set of data with a certain amount of slack. When Newton's theory is finally unseated, it is unseated, not by phenomena, but by a better theory. And, indeed, that's the only way a theory *can* be unseated. What it takes to overthrow a good explanation is a better explanation. Nothing less will do. This is the essence of the new empiricism. It gives us what the old orthodoxy actually could not—a good *empirical* reason for accepting unobservables. And this is true

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 1, January 1972

XII. ADJUSTERS AND SENSE-DATA

SAM COVAL AND D. D. TODD

MANY matters in philosophy come to a junction with a notion of conceptual dependence. We are interested in a rather particular application of this notion but find ourselves involved in aspects of the general question: are all cases which exhibit this unclear notion of secondariness or dependence *necessarily* so dependent? In particular our interest is in the purported dependency between "looks," "seems," "appears" (etc.) upon "is" and some of its implications for some form of phenomenalism.

Paul Grice argued in "The Causal Theory of Perception," that although statements such as "It looks red," may "imply" or depend upon denial or doubt, nevertheless the state of affairs associable with such a statement is separable from the conditions of its doubt or denial. This has important consequences since one argument sceptical of the sense-datist position is that L-Statements, ("looks," etc. statements), being so necessarily attached to or dependent upon or implicative of doubt or denial, could not serve in an analysis of is-statements, which are not so attached. What we do is to reinforce Grice at this point and to provide a hypothesis about the relations between is-statements and L-statements which could be useful to a sense datum theory in that it avoids certain kinds of dependency of "looks" upon "is."

But whereas Grice shows that "seems" does not have certain necessary relations with doubt or denial and is therefore free to be used in an "analysis" of "is," our approach is more subterranean. We try to show that the entire dependency of "seems" upon "is" rests upon a pragmatic basis, the nature of which, when displayed, enables us to see the nature of this dependency as harmless and indeed helpful to the phenomenalist, in that he is thereby enabled to pose certain leading questions.

I

We take as our point of departure a widely-used argument from John Austin which is usefully sceptical of the possibility of being able to state the relations between is- and L-statements in any way that is not inhibitory to the sense-datist, that is, in any way which does not exhibit the dependency of L-statements upon is-statements.

Austin comments, after pointing out some of the differences as regards various uses of "looks" in such cases as (a) "the moon looks no bigger than a sixpence" and (b) "they look like Europeans," that, ". . . contrary to what some philosophical theories seem to imply, the notion of being a so-and-so must be prior to that of being like a so-and-so."[1] We take Austin to be making the general point that the notion of being an *x* must be prior to that of looking *like* an *x*, *appearing* to be an *x*, *seeming* to be an *x*, *looking* *x*-ish or being like (an) *x* in any respect whatever.[2] Or, more generally yet, being an *x* is "prior," in a sense to be explained later, to any of the "adjusted" or qualified cases of *x*, such as not a real *x*, a dream-like, illusory, hallucinatory, false, unreal, or other likenesses of *x*: the standard case is prior to its deviations.

If this principle is true then the sense-datist is in trouble. The use to which such terms as "looks" must be put by the sense-datist is precisely to introduce notions which are meant to be "prior," not "posterior" to, that of simply *being x*, or *being* an *x*. But if the notion of being a so-and-so, the standard case, must be "prior" to that of being like or being any other adjusted case of a "so-and-so," then terms such as "looks," "seems," and "appears" cannot do what the sense-datist wants of them. Indeed they already do primarily what he doesn't want. An Austinian would say that the sense-datist is, by his need to use adjusters, committed to the standard "preceding" its adjustments but actually wants for his position the *converse* of this, viz., that in *some* sense the deviations or adjustments "precede," or at least are not preceded by, the standard case, and this is the source of his mistake. It is *the sense-datist fallacy*. We hope to provide a useful sense in which the adjustments are not preceded by their standard.

---

[1] J. L. Austin, *Sense and Sensibilia* (Oxford, 1963), p. 33.

[2] *Ibid.*, pp. 41ff.

II

To understand this Austinian argument we need to get clearer about the standard-adjuster mechanism. Naturally, this mechanism is best understood in terms of the purposes it serves. Chief among these are (1) to facilitate, in one fashion, interchange of information, (2) to allow our language means for dealing with the world's unpredictability and novelty while (3) minimizing and making full use of our vocabulary.

Consider the first. If, say a predicate, is to be productively usable by a speaker for a hearer there will have to be a standard use pre-fixed between the two to allow the term to function effectively for the hearer when invoked by the speaker. Otherwise the hearer will find himself constantly in need of advertising to the particular speaker's intentions, and/or the speaker will find himself constantly in need of explaining his words. Speakers and hearers both need to be privy to the convention, which means, for descriptive terms such "blue," that they both have to know of which sort of case "blue" is meant to be the linguistic correlate. But merely to have a standard case to which the convention would be fixed would not be enough. If we are to describe the world with a public language there will be two competing factors at work. There is our need as speakers and hearers for fixity in the language so that the intentions of the speaker are readily available to the hearer. There is, secondly, our need, if our language is to be adequate to the world it is meant to describe, for devices which allow us to deal not just with its uniformities, but with its novelties, and multiformities as well. We need two contrapuntal sets of devices, devices of fixity and devices of flexibility. Our standard-adjuster mechanism gives us both. Naturally, we also want these needs served with an economy and perspicuity of vocabulary.

With the standard-adjuster mechanism we are armed with a "flexibility-device by whose aid, in spite of the limited scope of our vocabulary, we can always avoid being left completely speechless."[3] We do not then need to strike a new term when faced with a new turn in the worlds, for with the aid of adjusters such as "like," we can bring to bear all of the riches of our already incorporated stock of terms. This allows us to defer the question of whether a new term ought to be struck or not. Without the deferment, allowed by "like"-plus-substantive, our vocabulary would be a great deal larger than it is. Witness such words as "dream-like," "coffee-flavored," etc. Clearly that is a saving in vocabulary; but the fact that there is such a saving is only a useful by-product; it is not an explanation of why the standard case retains its dominance over other cases such that they are treatable as its deviations. Why are decoy-ducks still known by the substantive? Why is an unintentional slight still called a "slight," false promises still "promises," illusory stars still "stars"? One can see that this last question is directly on the mark for the sense-datist and his opponent, who have different views on the relations between what is the case and what looks as if it is.

III

There probably won't be a comprehensive answer to this question of why the substantive is retained rather than untying the concept from its deviations since the various sets of adjusters—action adjusters such as "unintentional," "accidental" physical object adjusters such as "seems," "not real," "copy," "false"—will each have various relations with the standard from which they take their departures. But the following obvious things can be said. Decoy ducks look quite a lot like real ducks; they are after all modeled upon them; they may even fool us. That is why it is productive for us to say to the errant hunter that that's not a real duck. He has an excuse for taking the decoy for a duck, i.e., it is a "normal" error, and there is thus a point to correcting him in that way. He has not, however, the same excuse for taking that boat for a duck. Hence in this case we would not say that that's not a real duck. That correction is saved for peculiarly marginal cases. The phrase "not a real duck" implies that the case before us is a good deal like a duck, but misleadingly like one. Their intimate though deviant relations with the standard are good reasons for retaining the substantive "duck" for what we call "decoy-ducks." It is the role of the adjuster to specify these deviant relations and one of the roles of the substantive to provide the point from which such deviations are to be taken as departures.[4]

Given our usual preference for ducks over decoys and our interest in keeping such misleadingly similar cases separate but variously marked, *there is no more perspicuous or economical way of dealing with these*

---

[3] *Ibid.*, pp. 41ff.

[4] *Ibid.*, p. 74.

complex and subtle relations than with a standard-adjuster mechanism.

IV

The reasons for the existence of the *S-A* (Standard-plus-Adjuster) mechanism is one matter; the reasons for selecting which cases will fill the role of standard, and which the role of deviant, is another matter altogether.

In this connection it is crucial to note that it is in principle possible to have the value for *S* be the value for *A* and vice versa. This is *the principle of role-shift.* Less surprisingly it is possible to have the value of the adjusted case become the value for the standard case but not to require the complete inversion of roles where the previous value for the standard case becomes the value of the adjuster case. That is, converse or partially converse role-shift is possible.

Some role shifts would fill in as follows. A converse shift would be achieved when we took what we now call the pretense of scratching and made it the standard case because we found say certain overriding consequences now attaching to the case which we now call pretense and kept what is now actual scratching subserviant to our new standard because it was sometimes quite like it but never had the same important consequences. The present case of scratching no longer served, let us say, our dominant purposes but it did anyway still service our purposes to keep it conceptually linked to our new standard. Partial role-shift would be where we made what is now the pretense of scratching newly a standard case in its own right but didn't retain the present standard of scratching as its conceptual satellite, if indeed this is possible. Obviously, there are going to be other forms of relative interchange in the standard and adjuster roles of these two actions, scratching and pretending to. If we can make out the plausibility of any one type of shift our point would be served. That point is that we are able to separate off the conceptual mechanism of "standard and adjuster" from the cases to which that mechanism is applied.

There is, however, a serious impediment in the nature of some cases to that separation being made out. We have at least two main types of case before us exemplified in the difference between such cases as pretense and irony on the one hand, and let us say shadows on the other. It is the nature of the pretense of scratching that it cannot be adequately described without the mention of actual scratching. We could describe the behavior associated with the pretense and give this description independently of the concept of an actual scratch—he did thus and so —but if the intention is also a part of the pretense then we cannot describe that without the concept of the actuality since the intention of a pretender must be to produce an order of likeness to the actuality. If the intention of a pretender cannot be otherwise described it appears that a shift in role of the pretense out from under the dominance of the actuality or standard is not possible.

We must then set aside these cases which seem to exhibit a necessary secondariness. It is, however, natural that there be such recalcitrant cases since these are cases whose very existence is the result of a rationale which pre-suppose an antecedent standard. Irony is another example. What these necessarily secondary cases essentially show us is that after we have set ourselves certain conventions and/or institutional practices we only then have produced the possibility of various deviations from or conceptual reactions to these practices. These deviations therefore have as a necessary condition certain non-deviations in a way which does seem to run counter to the separation thesis.

Let us try the separation thesis on another case. The anti-separationist would hold that the concept of a shadow would entail that there was a physical object whose mass occluded the relevant light. A shadow he would say is always a shadow of something in such a way that a separationist would be required to deny an integral aspect of the concept, something in its meaning, to separate these two aspects—a "shadow"[5] and the fact that it is a "shadow"-of-something. It is pretty clear, however, that the dependency of the concept of a shadow upon the concept of an object will have to come from such factors as (1) that a shadow is always the result of something which happens to an object, (2) it doesn't always happen to an object with the result that the shadow is discontinuous, (3) its condition and/or shape depends very strictly upon such conditions so that it is characteristically amorphous, (4) shadows are two dimensional.

Perhaps we cannot imagine all of these not to be the case. Certainly we can imagine (2) and (3) not to be the case. If the separation thesis cannot be made out it will then have to be factors (1) and (4) which stand in the way.

---

[5] We propose to use "shadow" to mean the phenomenal shadow—that darkened patch which is the result of light's being occluded by an object. We will use "shadow" without quotes to mean what the anti-separationist would claim it to mean, viz., that if there is a "shadow" then it is entailed that an object is occluding light.

It cannot, however, be simply that since "shadows" are, by hypothesis, necessarily the result of something which happens to an object therefore "shadows" will be conceptually dependent upon objects. But even objects themselves can be the results of what happens to happen to another object or set of entities. If non-separation is to be even partly due to factor one it must be because of the special nature of the causal dependency of a "shadow" upon its object. But surely we could always imagine that to change: the special to become (the) ordinary. Suppose that what is unique in the dependency is that since shadows are normally rather near their objects we can mostly relate them one to the other. Imagine then that shadows are almost never near their objects and that, as a consequence we are only with great difficulty able to identify a shadow with its particular object since shadows are say only thrown at such a great distance. Or, imagine that although we know the general causal story of shadows and their objects we have as much difficulty discovering which particular object has cast a certain shadow as we have in discovering which particular automobile in traffic is identifiable as the source of which volume of exhaust. Of course, such identification can be achieved but if the difficulty is great and the gain small our concepts would reflect this and keep such things separate. In general, anyway, if the conceptual link between a shadow and its object is argued to be due to the special nature of their causal relation that would still mean that their conceptual link was reflective of a state of affairs which could be otherwise: we could find the causal relation altered. Such a conceptual linkage should consequently be breakable. That is, it is highly probable and certainly desirable that the language will not deny itself resources which allow such contingencies, however deep, however useful, however surrounded with practice, to be recognized and stateable as contingencies, albeit as deep contingencies. A language ought not deny itself such a resource since it then needlessly shuts off this type of legislation from the world. And even though legislations such as that a shadow will always be a "shadow" of something, are not in need of a continuing sensitivity to the world—that is after all why they are syntactically or permanently markable—nothing is gained by the desensitization of the philosopher to the source of such conceptual contingencies.

The other reason which the anti-separationist could have used was the two-dimensionality of shadows. One might argue that a dimensionality greater than two was required for an entity to qualify as a case of a physical object, that paradigm of standards. But that is not the issue. The issue is whether a dimensionality greater-than-two is required for an entity to qualify as a non-secondary phenomenon. There do seem to be cases, however, which clearly are not conceptually tied as adjuncts to other entities even in the *prima facie* way shadows are said to be tied by the anti-separationist and these cases are arguably of a dimensionality not greater-than-two. These cases are such things as rainbows and flashes. We are not in the case of rainbows forced on by a preposition such as "of" to include some other entities in our analysis of them. We can point to them and describe them as directly and simply as if they were bridges. It might be complained that rainbows are not truly two dimensional entities; but then neither are shadows, nor, literally, anything else for that matter.

The same conceptual independence holds for flashes. While it is true that we speak of the flashes of the guns, or a flash of lightning, this prepositional form is not necessary for the discussion of flashes: no more so than is a reference to the statues of Florence required to employ the name of that city or the name of any other particular position of their emplacement. While it is necessarily true that flashes are always caused and statues always emplaced and thus that in this sense every entity has some form of dependence upon others, we can assume that the anti-separationist is not making this point since it is at best irrelevant to his position. To make that point would be for him to say that all entities are in the above general sense dependent upon others. Irrelevantly, this is as true of bridges as it is of their shadows, of guns as of their flashes; this is the same sort of dependence for all, emanating as it does from the above causal and spatial truths which hold for both sets of entities. Therefore this kind of dependence of flashes upon objects and other phenomena is fully compatible with the separation thesis if indeed the anti-separationist is to allow that even physical objects such as bridges are themselves to be conceptually separable in the relevant sense.

The foregoing is intended to clarify what is meant by the claim that the case which plays a role of conceptual dependence can be seen as conceptually separable from that dependence. The case, not the role, can have its conceptual dependence cancelled under certain conditions. It is important to draw attention again to the distinction between a conceptual mechanism which is constituted by an interplay of a dependent and independent part, on the one hand, and the cases upon which that mechanism will work. What we have tried to show is that the concept of a "phenomenon which is the result of lights being occluded in certain ways such that this phenomenon is, etc." is necessarily a dependent concept because it cannot be analyzed without, or separated from, the concept of a physical object. The phenomenon itself, however, "the shadow," need not be seen as non-separable from its role within the concept. If a phenomenon can be so separated from its thus contingent role then although "seems" may in a sense not be separable from "is" the phenomena with which they engage may be separable. We now turn directly to this proposition.

Take the case of a blue shirt. Why and when do we say that the shirt is blue rather than it seems blue? We say it "seems blue" when there is reason to think that, for one thing, the light conditions may be misleading in some way affecting color perception or when there is reason to suspect that our visual apparatus is not up to the mark in some way affecting our perception of color. The fact that when we qualify our claim that the shirt is blue we must have reasons for such qualification implies that there is a circumscribable set of conditions under which we would *un*qualifiedly say that the shirt is blue.[6] These conditions are such as that kind of sunlight which allows us maximum visual discrimination at ordinary distances and that condition of the visual apparatus which allows us maximum visual discrimination[7] compatible with such discrimination being overwhelmingly common among perceivers. Some cases of color-blindness for instance allow less visual discrimination and are, as well, statistically rare, and thus fail on both counts. Cases where perceivers were capable of greater discrimination but which fell among the statistically rare would probably not serve well as the standard or unqualified case of perception because of the powerful utilities of there being a *common* set of perceptions among perceivers; but such a case might be allowed as a supplement to the standard. The practical possibility of communication about visual matters rests upon this common set of perceptions. It is to such common perceptions that our standard verbs of perception will therefore refer.

In short, the circumscribable conditions under which we would say simply that the shirt is blue are conditions which are productive of certain obvious goals such as coherence, discriminability, and communicability. Therefore those cases which best serve those goals will be the cases we select to play the role of the standard. The reasons we have related one set of cases to the sentence, "The shirt is blue," and another set to, "The shirt seems blue," are that we prize the greater set of coherent cases, the greater communicability, the greater degree of occurrence that the first set affords.

Put very generally then, our point is that the goals of coherence and communicability, and commonness of occurrence are among the criteria which determine which cases we shall select from the environment to ensconce as the standard. Were, however, the environment to change so that, say, ultraviolet light became the prevailing illumination, then our choice of which case was to serve as standard might also change. Furthermore, if we can imagine the remaining permutations of possible change in the above case, we can also imagine other resultant changes in the set of standard cases. Such contingencies logically permit the separation of the role which a certain case may play for us, from the case itself, or rather from the case playing some other role.

V

We return now to relate the above matters to the dispute between Austin and the sense-datist. Austin's persuasive suggestion, the consequences of which we have been dealing with, was that "the notion of being a so-and-so must be prior to that of being like a so-and-so." The question is whether the conceptual dependence or secondariness of "seems" to "is" is inimical to the sense-datist case.

The "priority" of "is" to any of its adjusters stems, we have argued, from our paramount need for a standard, fixed, case to serve the needs of communication. This need of a basepoint is uninstructive, however, as to the positioning of that basepoint. The need alone for standard cases does not help us to decide which cases will best fill that role. The need for the distinction between "looks" and "is" is in itself ontologically unselective. The position of that basepoint, or the actual selection of the cases which will serve as standard, is another matter; it is these other considerations as described in Sect. IV which decide ontological priorities. Thus, although "is" is prior to its adjusters in the sense outlined in Sect. III, it is not true that the cases selected as standard are necessarily "prior" to those cases which are treated as deviants.[8] Real coffee (the stuff) is not "prior" to ersatz coffee (the stuff) nor is the hallucinated dagger relevantly "posterior" to the real dagger. There is nothing about the world which makes the real dagger "prior," in Austin's sense, to the unreal one; the fact of there being an unreal one. We could untie one from the other instead of taking the conceptual advantage allowed us by keeping them tied as standard and deviant. We *could*, that is, introduce what we now call hallucinated daggers into our conceptual scheme independently of coffee. We still might want to say that they are respectively like a dagger and like coffee, but it would then be clear that they *need* not be conceptual satellites of coffee and daggers.

---

[6] See S. Coval and T. Forrest, "Which Word Wears the Trousers?", *Mind*, vol. 76 (1967), pp. 73–82.

[7] Which set of perceptions allows us a greater number of useful discriminations and which set perhaps a greater number, but few useful, discriminations and which set simply fewer discriminations is a collection of questions on which we can only satisfy the reader that we are aware of.

This separability is what Austin has overlooked in his otherwise true observation that "is" is prior to "looks." The sense-datist would allow and indeed embrace this conceptual priority of "is" over "looks." But this would not rule out the *rationale* of his question as contained, e.g., within the Argument from Illusion. That question is: what are the similarities between a particular standard case, (a case of is-*x*), and a deviant case, (a case of looks-*x*)? Instructively, the possibility of such a set of similarities itself implies a set of terms "simpler" than the terms being compared. We cannot, for instance, compare even one house with another except in terms of their features: *comparison is a three* term relation.[9] Comparison then between a standard and an adjusted case is just that sort of relation which harbors positive implications for the sense-datist position; *and the relation between "looks" and "is" already implies comparison.* It was, however, the point of Austin's observation to block such a comparison by elevating the standard case in such a way that it could not appear as a term in a comparison with a deviant case, since the moment that is allowed the sense-datist has indeed a question to ask. That question would be, as it has always been, "Mustn't we postulate some 'perceptual elements' in order to provide the third term in the similarity between veridical and non-veridical cases? And won't this third term be metaphysically neutral about the distinction between the veridical and the non-veridical since it is meant to be common to both?" We take this to be well along the road toward the introduction of sense data.

University of British Columbia AND
Simon Fraser University

Received March 9, 1971

---

[8] Another version of our problem might take the form of the complaint that it is impossible to introduce "sense data" via looks-*x* cases without using physical object terms (standard terms) as values for '*X*'. It has been answered to this by Benson Mates in "Sense Data," *Inquiry*, vol. 10 (1967), pp. 225–244 that this is merely the minor linguistic problem of indicating the extension of a new term via expressions which have established intensions. Such an answer, while correct as far as it goes, would ignore the logical dependence of the adjuster upon the standard, which we have been at pains to bring out and without which the argument cannot be made.

[9] See M. Mandelbaum, "Definiteness and Coherence in Sense Perception," *Nous*, vol. 1 (1967), pp. 123–138.

American Philosophical Quarterly
Volume 9, Number 1, January 1972

XIII. ENTHYMEMATIC CONSEQUENCE

ROLF GEORGE

I

ENTHYMEMES were traditionally defined as incomplete or incompletely stated *syllogisms.* Arguments of this sort, though formally invalid, must be allowed to have some merit, and although the restriction to syllogisms is undesirable, the definition at least has the advantage of precision. If the limitation to syllogisms is removed, notorious difficulties arise since a valid argument may be generated from even the most flagrant *non-sequitur* by the addition of a suitable premiss. Thus if we leave the domain of syllogisms, it seems no longer possible to give a formal distinction between formally invalid arguments that have some worth, and those that do not. In view of this some logicians, e.g., Messrs. Anderson and Belnap, to name only two, prefer to call all formally invalid arguments enthymematic.[1] Most others suggest that a token number of formal restrictions be placed upon the missing premiss (e.g., it must not contradict any of the given premisses, it must not be identical with the conclusion), but otherwise make the decision whether or not an invalid argument is enthymematic a matter of conscience and imagination.

I shall argue in this paper that, while it is *true* that enthymematic arguments can be augmented so that valid arguments result, it is not wise to *define* enthymemes in these terms. I shall instead give a definition of enthymematic consequence which is similar to Tarski's definition of logical consequence; one can even arrange matters so that the latter becomes a limiting case of the former. The definition can then be used to generate additional premisses which will convert enthymematic arguments into logically valid ones. It will thus automatically provide the desired restriction upon missing premisses.

I shall then show that the definition gives the same results as the traditional account within the domain of syllogisms, and that outside this domain it singles out a class of invalid but plausible arguments which seem to answer to many logicians' intuition of what an enthymeme is, if we can take the examples in their textbooks as a clue.

II

Following Tarski,[2] but in a somewhat less technical vein, we can define logical consequence as follows:

> The sentence X follows logically from the set of sentences K if, and only if, every substitution upon all extralogical constants of K and X which turns all the sentence of K into true sentences also turns X into a true sentence.

Analogously, I define *enthymematic consequence* in the following way:

> The sentence X follows enthymematically from the set of sentences K if, and only if, (*a*) X does not follow logically from the set K, (*b*) there exists a proper subset α of the extralogical constants of K and X, some of whose elements occur both in K and in X, and (*c*) every uniform substitution upon the constants in α which turns all the sentences of K into true sentences also turns X into a true sentence.

The first condition in the definiens reflects the general view that enthymemes are not formally valid arguments. The second condition is to forestall the consequence that a true sentence is enthymematically implied by any other sentence or set of sentences. I take both of these conditions to be uncontroversial. The third condition will be discussed in the sequel.

III

It is a major fault of the received account that it requires the missing premiss of an enthymeme to be true. This has the consequence that there are certain arguments of which we do not know whether they are enthymemes until we have come into the possession of a certain amount of factual information: "The Enterprise was commissioned in 1814, therefore the Enterprise was not a ship of the line," for example, requires the consultation of a naval almanac. I shall adopt the convention that any putative argument which fulfills conditions (*a*) and (*b*) above be called an enthymeme, and that any argument which also fulfills condition (*c*) be called a valid enthymeme. This makes it possible to characterize enthymemes formally. On the received view, "Socrates was a man, therefore Socrates was mortal" is considered a paradigmatic enthymeme, while "Jupiter is a god, therefore Jupiter is mortal" is not taken to be an enthymeme at all, even though both have the same form. I prefer, instead, to call every instance of the form "x is a ϕ, therefore x is ψ" an enthymeme, but to regard as valid only those instances of ϕ and ψ for which every substitution upon x which makes the premiss true also makes the conclusion true. Hence, "Socrates was a man, therefore Socrates was mortal" is valid in this sense. Thus the above definition allows us to make a tri-partition, and to distinguish valid and invalid enthymemes from grosser *non-sequiturs*.

---

[1] Alan Ross Anderson and Nuel D. Belnap, "Enthymemes," *The Journal of Philosophy*, vol. 58 (1961), pp. 713–723; they make the proviso that the conclusion be true.

[2] Alfred Tarski, "On the Concept of Logical Consequence," in *Logic, Semantics, Meta-Mathematics* (Oxford, 1956).

IV

I shall now show that for the domain of syllogisms the definition is equivalent to the traditional theory, i.e., that it will pick out the same truncated arguments as valid enthymemes. Consider "No shellfish are vertebrates, therefore some sea animals are not vertebrates." Since the addition of the further true premiss "Some sea animals are shellfish" results in a valid syllogism, it would be considered enthymematic in traditional logic. It is easy to show that this missing premiss is equivalent to the condition imposed by the definition of enthymematic consequence: The definition declares the argument enthymematically valid, because every substitution for "vertebrate" which results in a true premiss also results in a true conclusion. Using obvious abbreviations, we can state the condition which the definition requires, and which this argument satisfies as

(a) $(\forall\phi)\ [(\forall x)(Sx \supset \sim\phi x) \supset (\exists x)(Ax\ \&\ \sim\phi x)]$

i.e., no matter what we put for ϕ, whenever "$(\forall x)\ (Sx \supset \sim\phi x)$" is true, "$(\exists x)\ (Ax\ \&\ \sim\phi x)$" is also true. This condition, and analogous conditions for other syllogistic enthymemes, follow in each case from the so-called missing premiss, because, if S_1 is the missing premiss, and S_1, $S_2 \vdash S_3$ a valid syllogism, then it holds that $S_1 \vdash S_2 \supset S_3$ and consequently also that $S_1 \vdash (\forall\phi)\ (S_2 \supset S_3)$, where $(\forall\phi)\ (S_2 \supset S_3)$ is the universal closure over the predicate not contained in S_1.

In the present case it holds that

1. $(\exists x)\ (Ax\ \&\ Sx), (\forall x)\ (Sx \supset \sim Vx) \vdash (\exists x)\ (Ax\ \&\ \sim Vx)$

This is a valid syllogism. Thence

2. $(\exists x)\ (Ax\ \&\ Sx) \vdash (\forall x)\ (Sx \supset \sim Vx) \supset (\exists x)\ (Ax\ \&\ \sim Vx)$
3. $(\exists x)\ (Ax\ \&\ Sx) \vdash (\forall\phi)\ [(\forall x)\ (Sx \supset \sim\phi x) \supset (\exists x)\ (Ax\ \&\ \sim\phi x)]$

Conversely, the condition derived from the definition also entails the missing premiss. By instantiating on (a) with $\sim S$ for ϕ, we obtain

$$(\forall x)\ (Sx \supset \sim\sim Sx) \supset (\exists x)\ (Ax\ \&\ \sim\sim Sx)$$

By detaching from the tautology and double negation the missing premiss results. Unfortunately, I cannot provide a succinct general proof that the definitional condition always entails a categorical proposition which qualifies as missing premiss under the traditional theory. However, this result can be verified by testing every case, a procedure too tedious to bear repeating here. As an additional illustration I give the proof for the enthymeme B(a)roco: any enthymeme of this form is valid if, and only if

1. $(\forall\phi)\ [(\exists x)\ (\phi x\ \&\ \sim Mx) \supset (\exists x)\ (\phi x\ \&\ \sim Px)]$.

We obtain

2. $(\exists x)\ (Px\ \&\ \sim Mx) \supset (\exists x)\ (Px\ \&\ \sim Px)$ 1, U.I.
3. $\sim(\exists x)\ (Px\ \&\ \sim Mx)$ 2, MT
4. $(\forall x)\ (Px \supset Mx)$ 3, Dual

where (4) is the missing premiss.

V

The definition of enthymematic consequence also gives satisfactory results when applied to non-syllogistic enthymemes. For example, in the argument "Al is older than Bill, Bill is older than Charlie, therefore Al is older than Charlie," every substitution upon "Al," "Bill," and "Charlie" which makes the premisses true also makes the conclusion true; formally: $(\forall x)$ $(\forall y)$ $(\forall z)$

$[(Oxy \,\&\, Oyz) \supset Oxz]$, which is the statement of the transivity of "is older than." In general, all arguments which are based upon the properties of relations, but do not specifically state these properties in a premiss, will be acknowledged as valid enthymemes by the definition, and the condition required by the definition is identical with what is ordinarily given out as the missing premiss in these cases. Hence the definition handles, in addition to syllogistic enthymemes, also the numerous enthymemes which rely on the formal characteristics of relation predicates. Moreover, I have not been able to find any intuitive enthymemes which were not warranted under the definition, though certain cases yielded only after some analysis. For example, "It rains, therefore the streets are wet" is clearly a meritorious argument which we are inclined to consider enthymematic, but it does not contain any term which premiss and conclusion have in common. It appears that cases of this sort prove that enthymemes must be defined in terms of missing premisses: The additional premiss, "If it rains, then the streets are wet" generates a valid argument; it is also obviously true. Hence it seems that our theory fails, while the missing premiss doctrine applies. But this is somewhat misleading: to the extent in which the missing premiss is obvious, it must be considered tacitly quantified, i.e., we must not take the antecedent and consequent to be atemporal, but must assume that what is meant is "if it rains, then the streets are wet *at that time*," or, "for any time t, if it rains at t, then the streets are wet at t." Similarly, the original argument should be presented as "It rains *now* (or at t_1), therefore the streets are wet now (or at t_1)." Hence in this case the common term required by the definition is the time constant. This will often be the case when the missing premiss is a law-like sentence. In other cases it will be necessary to analyze the expressions occurring in the argument, because an unanalyzed enthymeme often does not show forth the terms which make the definition applicable.

Since I have found no conspicuous exceptions, I assume that the theory here propounded explicates the notion of an enthymeme successfully, not only for syllogisms and relational arguments, but for formally defective inferences *schlechthin*.

VI

I said above that it is true that an enthymeme can be converted into a formally valid argument by the addition of a further premiss. It should be obvious by now that the definition of enthymematic validity allows us to generate such a premiss mechanically: it is the universal closure over the "crucial" term(s) of the conditional formed by the conjunction of the premisses and the conclusion. If this proposition is true, i.e., if the conditions of the definition are satisfied, the propounder of a valid enthymeme can insist that he has done just this: propounded a valid enthymeme. He has done a respectable job; he no longer has to rely on the benevolence and charity of his audience, as Rescher suggests he must.[3]

On the view here put forth, enthymemes are a term-logical phenomenon; the definition disqualifies as missing premisses of an enthymeme the material conditional consisting of the conjunction of premisses as antecedent and the conclusion as consequent, and thereby disqualifies a very large number of troublesome *non-sequiturs* which heretofore seemed to make appeals to obviousness etc. of the missing premisses necessary.

VII

I now wish to show that in some cases the definition gives very clear and obvious results where the traditional view might lead to controversy concerning the nature of the missing premisses. Consider "Sleipnir is a stallion, therefore the head of Sleipnir is the head of a horse." Every substitution upon "Sleipnir" which makes the premiss true also makes the conclusion true. Hence the argument is a valid enthymeme under the definition; but what are the missing premisses? It is not my aim, in the next few paragraphs, to determine these premisses. On the contrary (setting aside the result of the preceding section), I want to raise a bit of a controversy on this score to show that one can be quite certain that a certain argument is a valid enthymeme without being certain about the nature and number of the missing premisses. One of them, namely, "All stallions are horses" can be readily made out. But is it the only missing premiss? It seems that there are two further, rather more doubtful, candidates.

First, it appears that the conclusion entails "Sleipnir has a head." But this proposition is not entailed either by "Sleipnir is a stallion," nor by the already added premiss "All stallions are horses." Hence a further premiss seems to be required, namely either "All horses have heads" or "All stallions have heads," a proposition to which most will assent.[4]

---

[3] Nicholas Rescher, *Introduction to Logic* (New York, 1964), p. 162.

But it seems that there is a further missing premiss. Unfortunately, our language does not as yet contain any means of expressing it. It is best exhibited by way of an analogy with another enthymeme, namely, the argument based upon the transitivity of "is older than" which was stated earlier. Professor Copi, quite reasonably, said about that enthymeme that the missing premiss asserts that "is older than" is a transitive relation.[5] It is clear that the relation must satisfy certain conditions to sustain the argument. If we replace "is older than" in the earlier argument by "votes for," the result is an invalid enthymeme. A similar point can also be made about the relation indicated by "is" in "Sleipnir is a stallion": if we replace "is" by "looks like," we obtain the invalid enthymeme "Sleipnir looks like a stallion, therefore the head of Sleipnir looks like the head of a horse." This is invalid because there are some stallions whose heads do not look like the head of a horse. For example, the substitution of "Bucephalus" in this last argument results in a true premiss but false conclusion, since, as Montaigne observed, "Every one knows, that Alexander's horse Bucephalus had a head inclining to the shape of that of a Bull." But this means that the relation indicated by "is" supports an argument which the relation indicated by "looks like" does not support. On the received doctrine, this fact should be stated in an additional premiss. But since we have no general expression for relations of this sort, it shall here remain unexpressed for lack of the requisite vocabulary.

This section was to show that we are often quick in recognizing a valid enthymeme, but slow in making out what the additional premisses might be that would turn them into formally valid arguments.

VIII

It remains to give a logician his due who more than a hundred years ago propounded a theory of logical consequence which in one definition accounted for both logical and enthymematic validity: Bernard Bolzano. He defined consequence thus: "I say that propositions M, N, O . . . follow from Propositions A, B, C, D, . . . with respect to the variable parts i, j, . . . if every class of representations whose substitution for i, j, . . . makes all of A, B, C, D, . . . true also makes all of M, N, O, . . . true."[6] It has been pointed out that Bolzano anticipated Tarski by almost exactly a hundred years in his definition of logical consequence.[7] Indeed, if the variable parts i, j, . . . are taken to consist of all and only the extralogical terms of A, B, C, D, . . . , M, N, O, . . . the definition is close to that of Tarski (though Bolzano demands that the premisses be consistent). Cases where i, j, . . . include more or fewer than the extralogical terms were generally regarded as somewhat quaint. In particular, it has not been seen that cases of Bolzano-entailment where the class of "variable" terms is smaller than the class of extralogical terms are just those argument forms which we are wont to call enthymemes.[8]

University of Waterloo

Received August 21, 1970

---

[4] Though not all. M. Oldfield Howey writes: "Sir Thomas Boleyn (the father of Anne, R.G.) is doomed for a period of a thousand years annually on a certain night, to drive a coach which is drawn by four *headless* horses in a circuit which takes him over twelve bridges in the neighbourhood . . . On these occasions Sir Thomas carries his head under his arm whilst flames come forth from his mouth." M. Oldfield Howey, *The Horse*, New York (Castle Books), no year. Emphasis added. Howey would have to consider the enthymeme invalid.

[5] I. Copi, *Symbolic Logic* (New York 1967), p. 155.

[6] Bernard Bolzano, *Wissenschaftslehre*, Vol. 2 (Sulzbach, 1837), p. 114.

[7] E.g., by Heinrich Scholz.

[8] This paper was written during a summer I spent at the Department of Psychology of the University of Alberta in Edmonton, upon the invitation of Prof. Thomas M. Nelson.

BOOKS RECEIVED

Edwards, Paul, *Buber and Buberism: A Critical Evaluation* (Kansas: University of Kansas [The Lindley Lecture], 1969), 49 pp.

Fakhry, Majid, *A History of Islamic Philosophy* (New York: Columbia University Press, 1970), 427 pp.

George, F. H., *Computers, Science and Society* (New York: Prometheus Books, 1970), 101 pp.

Hilpinen, Risto (ed.), *Deontic Logic: Introductory and Systematic Readings* (Holland: D. Reidel Publishing Co., 1971), 182 pp.

Kurtz, Paul (ed.), *Language and Human Nature* (Missouri: Warren H. Green, Inc., 1971), 251 pp.

Kraft, Victor, *Mathematik, Logik und Erfahrung* (New York: Springer-Verlag, 1970), 104 pp.

Kurtz, Paul and Stojanovic, Svetozar (ed.'s), *Tolerance and Revolution* (New York: Prometheus Books, 1970), 165 pp.

Lal, R. B., *The Gita in the Light of Modern Science* (Bombay: Somaiya Publications Ltd., 1970), 315 pp.

Lecaldano, Eugenio, *Le Analisi del Linguaggio Morale* (Rome: Edizioni dell'Ateneo, 1970), 283 pp.

Levensky, Mark (ed.), *Human Factual Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1971), 218 pp.

Mayeroff, Milton, *On Caring* (New York: Harper & Row Pub., 1971), 63 pp.

McLean, George F. and Aspell, Patrick J., *Ancient Western Philosophy: The Hellenic Emergence* (New York: Appleton/Century/Crofts, 1971), 332 pp.

Osborn, Reuben, *Humanism and Moral Theory* (New York: Prometheus Books, 1970), 115 pp.

PRINTED IN ENGLAND
by C. Tinling & Co. Ltd., Liverpool, London and Prescot

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, $6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, $6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, $6.00.

No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, $6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Robert Butts
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Arthur C. Danto
James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
R. M. Hare
Rom Harré
Carl G. Hempel
John Hospers
Raymond Klibansky
Hugues Leblanc
Ernan McMullin
Gerald J. Massey
Richard Rorty
George A. Schrader
Wilfrid Sellars
Alexander Sesonske
Manley H. Thompson, Jr.
John W. Yolton

VOLUME 9/NUMBER 2 APRIL, 1972

CONTENTS

I. MICHAEL J. LOUX: *Recent Work in Ontology* 119

II. HECTOR-NERI CASTAÑEDA: *Intentions and Intending* . . . 139

III. O. R. JONES: *After-Images* . . 150

IV. PATRICK SHERRY: *Is Religion a "Form of Life"?* . . . 159

V. A. HYSLOP and F. C. JACKSON: *The Analogical Inference to Other Minds* 168

VI. JACK NELSON: *Logically Necessary and Sufficient Conditions for Identity Through Time.* 177

VII. IAN HACKING: *The Logic of Pascal's Wager* 186

VIII. STANLEY L. PAULSON: *Two Types of Motive Explanation* . . . 193

IX. FRANK SNARE: *The Concept of Property* 200

X. PETER MILLER: *What Aristotle Should Have Said: An Experiment in Metaphysics* 207

Books Received 213

PUBLISHED IN OXFORD BY BASIL BLACKWELL

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost $12.50, increasing by $1.20 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost $10.10, increasing by $1.20 for each additional 4 pages. Covers will be provided for offprints at a cost of $6 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 2, April 1972

I. RECENT WORK IN ONTOLOGY

MICHAEL J. LOUX

IN this paper, I want to examine recent work (1959 to the present) on two central problem-areas in ontology: (1) problems about the existence and nature of universals and (2) problems about the structure and individuation of particulars. A survey of this sort can emphasize either breadth or detail; one can deal with a large number of papers and books in a general way, or he can single out certain important writings and discuss them at greater length. Both approaches have their advantages and disadvantages. In the present survey, I have tended to stress detail. To compensate for lack of breadth, I have supplemented the survey with a fairly complete set of footnotes where I cite writings I do not discuss in the body of the paper. Doubtless, some would have preferred a comprehensive bibliography; but while helpful, bibliographies frequently have the effect of confusing the reader. The difficulty is that titles, by themselves, often fail to give sufficient information about the content of a paper or book.[1] By relating them to the points made in the survey itself, I have tried to place unexamined materials in a definite context. Hopefully, this will provide the reader with a principle of selection, should he wish to pursue the issues discussed here.

I. THE PROBLEM OF UNIVERSALS

Recent literature on the problem of universals does not exhibit the sort of unity which frequently characterizes work on philosophical problems. While there is a certain minimal give and take among philosophers working in this area, there are no on-going debates dominating the journals and, consequently, no common starting-points for investigation. The general picture is rather one of individual philosophers working in relative isolation from one another, each starting from scratch, rethinking the very context in which the problem is to be approached.

All of this makes it impossible to provide a unified picture of recent work on universals; however, something less ambitious and less tidy is possible. One can talk about individual papers and books and point, in a general way, to similarities and differences of approach. This I shall try to do. I shall first examine the work of philosophers who can reasonably be construed as defending some version of metaphysical realism; then, under the slightly misleading heading, "Nominalism," I shall deal with the writings of philosophers who defend some other solution to the problem of universals.[2]

A. *Metaphysical Realism*

Throughout the history of philosophy, the analysis of predication has played a prominent role in the theories of metaphysical realists; nor is this surprising, as realists have often contended that it is in predication that universals are introduced into discourse. During the past decade or so, a number of philosophers have echoed this traditional theme. Very generally, their claim has been that predicates, along with subjects, must be construed as having referential force.[3]

Probably the most persistent exponent of this line of thought has been Gustav Bergmann. For more than two decades, Bergmann has defended the view that predicates are, in a fairly straightforward way, names of universals.[4] Utilizing the machinery of his Ideal Language approach to philosophy, he has argued that subject-predicate sentences like "This is red" involve entities of distinct logical types: the individual named by the subject-term and the common character named by the predicate-term. According to Bergmann, the point of such a sentence is to indicate that the named entities stand in the nexus of exemplification. This nexus is assymetrical. Individuals can exemplify characters, but they cannot themselves be exemplified. While he grants that exemplification is assymetrical, Bergmann contends that it cannot be a relation in any ordinary sense. Relations are themselves universals which objects can exemplify. Were exemplification a relation, one thing could exemplify another only in virtue of some higher level form of exemplification; and this new form of exemplification would require a still higher level form of exemplification, and so on *ad infinitum.*[5]

---

[1] For bibliographies on the issues covered by this survey, see Guido Küng, *Ontology and the Logistic Analysis of Language*, tr. by E. Mays (Dordrecht, Holland, 1967) and Michael Loux (ed.), *Universals and Particulars* (New York, 1970).

[2] A general account of the problem of universals is presented in Loux's "The Problem of Universals" in *Universals and Particulars, op. cit.* Kung's *Ontology and the Logistic Analysis of Language, op. cit.*, provides an excellent account of 20th century work on the problem. For accounts of classical positions on the problem, see Richard Aaron, *The Theory of Universals*, 2nd ed. (London, 1967); A. D. Woozley, "Universals," *Encyclopedia of Philosophy*, vol. VIII, pp. 194–206; G. E. M. Anscombe and P. T. Geach, *Three Philosophers* (Oxford, 1963); and Farhang Zabeeh, *Universals* (The Hague, 1966).

[3] For a discussion of issues central to this sort of approach see Rolf Eberle, "Universals as Designata of Predicates," *American Philosophical Quarterly*, vol. 6 (1969), pp. 151–157.

While Bergmann has remained constant on these points, his more recent writings have tended in the direction of an even more extreme Platonism. In his early writings, notions like "individual," "character," and "exemplification" are treated as meta-linguistic concepts; but in "Ineffability, Ontology, and Method" and subsequent papers, Bergmann contends that these concepts have an ontological status.[6] In this paper, he reiterates the claim that in being presented with a phenomenological object, like a red spot, I am presented with two things, an individual (*a*) and the simple character it exemplifies, (*F*). Bergmann makes the further claim, however, that in being presented with the red spot, I am presented with three additional entities (not "things")—the individuality of *a*, the universality of *F*, and the nexus of exemplification tying *a* and *F* together. In supporting this claim, he appeals to the Principle of Acquaintance. In knowing that the spot is red (that *a* is *F*), Bergmann claims to know that *a* is an individual, that *F* is a universal, and that *a* and *F* stand in the exemplification nexus; but he claims to be unable to account for his knowledge of these facts unless he grants that he is presentationally acquainted with the respective entities, individuality, universality, and exemplification.

Still, Bergmann denies that the ontological status of these notions is on a par with the status of individuals and the simple characters they exemplify. The latter are things; they exist. The former, however, are merely entities; they can only be said to subsist. Bergmann gives substance to these rather cryptic remarks by telling us that individuality, universality, and exemplification are ineffable; they cannot be named without (1) disaster and (2) futility. Apropos of (1), Bergmann argues that the attempt to name these entities generates a vicious infinite regress. If one is to name individuality, universality, and exemplification, he must construe these as the nominata of predicative expressions. Call these expressions '*I*', '*U*', and '*E*'. Now, Bergmann challenges us to use these expressions to state the facts at issue, i.e., that *a* is an individual, that *F* is a universal, and that *a* exemplifies *F*. Focusing on the first point, we write (in the symbolism of the Ideal Language), "*I*(*a*)." We may think that in this sentence the individuality of *a* is named; but is it really? No, Bergmann argues; it presents itself once again unnamed in the fact that '*a*', an individual descriptive sign, is coupled with a predicative expression. By introducing a new predicate into the Ideal Language, we might try to label the individuality of *a* as it appears in this sentence. Unfortunately, this will not remove the difficulty; for we can couple this new predicate with '*a*' to form a sentence in which the individuality of *a* once again emerges unnamed. No matter how we increase our stock of names, we are never able to exclude this possibility. The individuality of *a* is systematically elusive, and the same is true of the universality of *F* and the exemplification tying *a* and *F* together.

---

[4] See, for example, ch. IV, X, XI, XII, and XIII of *Meaning and Existence* (Madison, 1960). Also see ch. II, VI, VII, VIII, and IX of *Logic and Reality* (Madison, 1964) and Pt. I of *Realism: A Critique of Brentano and Meinong* (Madison, 1967). For attempts to clarify and defend Bergmann's position on the problem of universals, see Edwin Allaire, "Existence, Independence, and Universals," *The Review of Metaphysics*, vol. 14 (1960), pp. 485–496 and Reinhardt Grossman, "Conceptualism," *The Review of Metaphysics*, vol. 14 (1960), pp. 243–254. These and other papers having a Bergmannian flavor are collected in *Essays in Ontology* (The Hague, 1963).

[5] For a critical examination of this kind of argument, see Nicholas Wolterstorff, "Objections to Predicative Relations," *American Philosophical Quarterly*, vol. 7 (1970), pp. 238–245.

[6] "Ineffability, Ontology, and Method," *The Philosophical Review*, vol. 69 (1960), pp. 18–40; the paper reappeared as ch. II of *Logic and Reality, op. cit.* In this paper Bergmann wants to make the further claim that certain logical notions, such as the truth-functional connectives, have an ontological status. In "Existence and Generality," *Theoria*, vol. 28 (1962), pp. 1–26 later ch. III of *Logic and Reality, op. cit.*, he argues that what is expressed by the universal quantifier, as well as what is expressed by the existential quantifier, subsists. An interesting discussion of Bergmann's views on subsistence and ineffability is found in John Peterson, "Bergmann's Hidden Essences," *The Review of Metaphysics*, vol. 22 (1969), pp. 640–657.

As regards (2), the futility of naming these subsistents, Bergmann points to the fact that we can understand a language only when we know how the symbols of the language may be concatenated with each other, how symbols of different styles range over things of different types, etc. But to know these things is just to know the truths expressed by Ideal Language sentences like "$U(F)$" and "$I(a)$." It is to know, for example, that a symbol like '*F*' names an entity that is a universal and that a sign like '*a*' names an entity that is an individual. Thus, nothing would be accomplished by pinning names on subsistents like individuality and universality; in order that a person understand the sentences in which these names would appear, he would have to know in advance that the sentences are true.

It is difficult to assess Bergmann's work. This is especially true of his recent remarks about subsistence and ineffability. They are so deeply embedded in his Ideal Language approach that they appear immune to external criticism; nor do they invite internal criticism; for such criticism would seem to presuppose more agreement with Bergmann on methodological issues than one is likely to find among most philosophers. Nonetheless, it seems that one can quarrel (from a methodologically neutral stand-point) with the guiding principle of Bergmann's account, his view that predicates function as names. Bergmann's choice of examples makes this view appear plausible. He takes sentences like "This is red" and "This is green" as paradigms of the subject-predicate form; and in the case of these sentences, the view that predicates are names is not obviously untenable. If we turn to other sentences, however, the view is less appealing. In the case of "John is courageous," there is no temptation whatever to construe "courageous" as naming courage. The noun-form, "courage," plays that role. Thus, with "courage" and "courageous," we distinguish between name and predicate. Color words, however, are misleading. As Wilfrid Sellars points out in "Naming and Saying," an expression like "red" is three grammatically distinct words wrapped up into one—a singular noun, a general noun, and an adjective.[7] Qua singular noun (and possibly qua general noun), "red" can plausibly be construed as a name; but qua adjectival form, it cannot. Unfortunately for Bergmann, it is precisely qua adjectival form that 'red' appears in "This is red."

In "Universals and Metaphysical Realism," Alan Donagan presents what appears to be a different version of the view that predicates are referential in function.[8] He wants to claim that all primitive predicates occurring non-redundantly in true propositions denote universals. Donagan calls this claim the Realist Principle, and he attributes it to Russell and Moore. He never defends the principle in any detail; but he claims that a defense would involve showing (1) that predicative expressions cannot be analyzed into non-predicative expressions and (2) that the truth-value of propositions depends upon their relation to the world rather than upon their relation to the thinking of any individual.

Donagan's paper is extremely valuable in that he attempts to dispose of a large number of classical and contemporary criticisms of metaphysical realism. Nonetheless, the force of his own version of realism remains unclear; for he fails to explicate the notion of denotation that is operative in his formulation of the Realist Principle. Presumably, he means to use the term in the way that Russell and Moore did; unfortunately, neither of these philosophers was ever very clear about the notion. Donagan tells us that subject-expressions also denote "real things"; and he seems to hold that "denotes real things" is univocal over subjects and predicates. But, then, is Donagan's notion of denotation so very general that while "denotes" means the same thing for subjects and predicates, terms functioning as subject and terms functioning as predicate can, nonetheless, differ in style or mode of denotation? Or is Donagan committed to the view (held by Bergmann) that denotation is simply denotation, that subjects and predicates denote in precisely the same way? Donagan's paper raises, but does not answer, these questions.

P. F. Strawson's discussion in *Individuals* is much clearer on these points.[9] Strawson holds that, strictly speaking, predicates have no referential force. It is by means of subjects that we refer to things; by means of predicates, we predicate other items of these things. Thus, both subject and predicate introduce non-linguistic items (terms, in Strawson's terminology), but the style of introduction differs in the two cases. Semantically, we mark this difference by distinguishing between reference and predication; but even at the grammatical level, the difference comes through. Subjects introduce their terms in the substantival or noun-like style—in the style appropriate to lists; whereas predicates introduce their terms in the assertive or verbal style: inasmuch as predicates incorporate some form of the verb in the indicative mood, they carry the mark of propositional assertion.

---

[7] "Naming and Saying," ch. VIII of Sellars' *Science, Perception, and Reality* (London, 1963).

[8] "Universals and Metaphysical Realism," *Monist*, vol. 47 (1963), pp. 211–246.

[9] See Part II of *Individuals* (London, 1959). For a criticism of Strawson's account, see ch. V of John Searle, *Speech Acts* (London, 1969). In this same chapter Searle briefly argues that realism is true, but he contends that the truth of the view is essentially tautological. For a criticism of Searle's own version of realism, see Gerald Vision, "Searle on The Nature of Universals," *Analysis*, vol. 20 (1970), pp. 155–160.

Strawson defends a traditional view when he argues that the particular-universal distinction is intimately related to the subject-predicate distinction. It is not that universals can never be introduced by subjects; clearly, they can. The point of the traditional doctrine, Strawson contends, is rather that particulars are the paradigmatic referents of subjects. While, they are the subjects of the most basic kind of subject-predicate propositions, particulars can never be predicated of things of other kinds. Thus, in the basic kind of subject-predicate proposition, we couple a particular, as introduced in the substantival style, with a universal, as introduced in the assertive style. The predicated universal, however, can be of either of two types. It can be a sortal universal like *man* or *dog* or a characterizing universal like *green* or *courage*. In both cases, the universal collects all of those particulars of which it can be truly predicated; but the principle of collection differs. Roughly, a sortal universal collects all of those particulars which are instances of it; whereas, a characterizing universal collects all and only those particulars which are characterized by it.

An interesting feature of Strawson's account is his adaptation of Frege's views about saturated and unsaturated elements. Strawson's particulars, like Frege's objects, are saturated or complete; they unfold, he tells us, into facts. His universals, on the other hand, are like Frege's concepts in being incomplete. The point is that a particular can be introduced into discourse by a speaker only if he knows an individuating fact about that particular —only if he knows the truth of some empirical proposition that uniquely identifies the particular in question. In the case of universals, however, no such empirical knowledge is presupposed. All that is presupposed is the speaker's ability to use certain words correctly.

In a paper originally published in 1954 ("Form and Existence"), P. T. Geach expresses a view similar, in certain respects, to Strawson's; and although he is now less confident of the conclusions reached in that paper, Geach incorporates a revised version of "Form and Existence" in *God and the Soul.*[10] Like Strawson, Geach contends that subjects and predicates differ in mode of signification. Geach expresses this difference by saying that a predicate is true or false of what a subject stands for. While he finds this difference important, Geach thinks it legitimate to speak of the referents of predicates. He is never very clear about the nature of these entities, but he seems to construe them along the lines of Aquinas' forms and Frege's concepts.

Geach's account departs from Strawson's, however, on the subject of abstract singular terms. His view is that we cannot refer to the referents of predicates by using simple expressions like "wisdom" and "courage." Expressions of this sort purport to signify a "one over many"; but following the perplexing terminology of Aquinas and Frege, Geach claims that, in themselves, the referents of predicates are neither one nor many, although they can be "indifferently one or many."

As regards the use of abstract singular terms, Geach is avowedly nominalistic. He maintains that simple expressions like "wisdom" and "courage" are useful in that they provide handy abbreviations for longer, more complicated expressions; but he claims that philosophers have the right to use such terms only to the extent that they can eliminate them in favor of concrete forms like "wise" and "courageous." In this context, Geach goes so far as to say that a sentence containing an unanalyzable abstract singular term is unintelligible.

Geach is willing to grant, however, that we can refer to the referents of predicates by means of subjects. Such reference occurs when we speak of a form as individualized, when we use a complex expression like "wisdom of Socrates." Using an expression of this sort, we avoid the Platonistic mistake of construing a predicate-referent as a "one over many." We speak instead of a form as present in this or that individual. All of this is fairly straightforward. Geach goes on, however, to make some puzzling remarks. He says that in the complex expression, "wisdom of Socrates," it is not "wisdom," but "wisdom of" that stands for the form itself (the form as not individualized). "Wisdom of Socrates," he argues, does not break up into "wisdom" and "of Socrates." This would be to grant too much to Platonism. Presumably, Geach is attempting to defend the Fregean view that predicate-referents are incomplete, while maintaining, against Frege, that such things can be the referents of subjects. The difficulty here is that "wisdom of" is not a referring expression, but a fragment of a referring expression. Geach, I think, wants to have it both ways.

---

[10] "Form and Existence" originally appeared in *Proceedings of the Aristotelian Society*, vol. 55 (1954), pp. 251–272. The revised version appears in *God and The Soul* (London, 1969). A similar account is presented in "Aquinas" in *Three Philosophers, op. cit.* For a statement of Geach's recent doubts about the position, see "What Actually Exists," *Proceedings of the Aristotelian Society*, Supplementary Volume XLII (1968), pp. 7—16, reprinted in *God and The Soul*. For a response to this paper, see R. H. Stoothoff, "What Actually Exists," *Proceedings of the Aristotelian Society*, Supplementary Volume XLII (1968), pp. 17—30.

Of the philosophers we have so far considered, only Strawson talks of non-predicable universals. For Bergmann, Donagan, and Geach, the notion of a non-predicable universal would seem to be incoherent. According to each of these philosophers, universals just *are* the sorts of things named, denoted, or referred to by predicative expressions. Strawson does make a few remarks about non-predicable universals, but nothing like a general account of such entities emerges from his discussion. In "On the Nature of Universals," Nicholas Wolterstorff provides a more extensive treatment of this notion.[11] For Wolterstorff, non-predicable universals are things like literary works, films, musical works, and art prints. His claim is that things of this sort exhibit a one-many relation that is analogous to the one-many relation exhibited by predicable universals (things like properties and actions) and their instances. Thus, a literary piece can have several copies, several recitations; a film, several showings; a musical work, several performances; and an art print, several impressions. According to Wolterstorff, the analogy rests on the fact that non-predicable universals, like predicable universals, are kinds; qua kinds, they can have instances. All universals, then, are kinds of things; but, Wolterstorff contends, kinds are not to be identified with the classes of the logician or the mathematician. Identity-conditions for classes can be specified in purely extensional terms; not so for kinds: a particular kind could have had other members than those it does, in fact, have; and two kinds with precisely the same members are not necessarily one and the same kind.

In both Strawson and Wolterstorff, then, some universals cannot be introduced by predicates. In *Modes of Referring and the Problem of Universals*, David Shwayder defends the more extreme view that predicates never introduce universals.[12] According to Shwayder, the case for metaphysical realism rests not on the subject-predicate distinction, but on the fact that language incorporates abstract singular terms with genuine referring uses. Shwayder's defense of realism is, by his own admission, very restricted. He considers only properties; and as he defines the concept, properties are limited to things like colors and shapes. Shwayder argues that property-words have their primary use in predication; but their use in predication, he contends, carries no ontic commitment. Functioning as predicates, property-words are not used to refer to or denote properties, but rather to distinguish or characterize the objects referred to by subjects. Nonetheless, the use of property-words in predication provides a basis for secondary uses of these words. In their secondary uses, property-words function as abstract singular terms and carry referential force. Their referents, of course, are properties; but, according to Shwayder, a property just is the primary use of a property-word—its use in distinguishing and characterizing objects.[13]

At first glance, it is difficult to see how this constitutes a defense of what has traditionally been called realism. On Shwayder's view, properties or universals turn out to be mere uses of words. They are hardly the sorts of things traditional realists have so vigorously defended against the attacks of nominalists. Indeed, it would be difficult to find a nominalist who would quarrel with the claim that properties, in Shwayder's sense, exist.

In fairness to Shwayder, there are contexts where his view is presented in more metaphysically charged terms. Thus, he sometimes claims that in their secondary uses, property-words take as their referents not the distinguishing uses of predicate-terms, but rather the theoretically possible distinctions we could draw by the primary uses of property-words. Further, Shwayder contends that the introduction of secondary uses of property-words involves the introduction of a whole new department of language, including new kinds of criteria for individuation, identity, etc.

---

[11] Wolterstorff's paper originally appeared in Loux's *Universals and Particulars*, *op. cit.* It is part of a book, *On Universals*, which is to be published by the The University of Chicago Press (forthcoming).

[12] *Modes of Referring and the Problem of Universals* (Berkeley, 1961).

[13] In the interests of space, I oversimplify a bit here. Shwayder does not claim that for every primary use and its corresponding secondary use, there is a single word playing both roles. This does happen with words like "red"; but Shwayder realizes that in the case of expressions like "wise" and "wisdom," we distinguish between word playing predicative role and word playing referential role. What Shwayder is anxious to show is that even in these cases, the predicative use is primary and lays the foundation for the referential use.

A rather different approach characterizes Panayot Butchvarov's *Resemblance and Identity*.[14] According to Butchvarov, the analysis of abstract terms, as well as the analysis of predication, is irrelevant to the traditional problem of universals. What is relevant is the "recurrence of qualities." It is, Butchvarov contends, a fact of common-sense that individuals can agree in their qualities; the problem of universals has as its focus the appropriate philosophical description of this theoretically neutral fact. Most of Butchvarov's book is concerned with an examination of the forms of description provided by what he calls the Resemblance Theory and the Identity Theory. On the Resemblance Theory, to say that a quality, *F*, recurs in individuals, *a . . . n*, is to say that *a . . . n* possess resembling, yet distinct, particular qualities. The Identity-Theory, on the other hand, speaks of a single quality as numerically identical in different individuals.

Butchvarov presents an extremely complex argument against the Resemblance Theory. Roughly, his contention is that since resemblance between qualities lacks the formal properties it would have to exhibit to constitute a relation between numerically different individuals, it must be construed as a form of identity. In support of this claim, he argues that propositions of the form "*x* resembles *y*" are necessarily incomplete. It is not that a proposition of this sort fails to specify how the relevant objects are alike, but rather that it fails to specify the degree to which they are alike. Butchvarov contends that resemblance, if a relation at all, is a relation that "admits of extremely wide variation of degrees." Consequently to make the claim "*x* resembles *y*" intelligible, one must indicate the extent of the resemblance. The only way of doing this, he claims, is to compare the resemblance holding between our original objects with the resemblance holding between other objects. Thus, "*x* resembles *y*," if intelligible at all, is short for "*x* resembles *y* more than *w* resembles *z*." In this respect, "*x* resembles *y*" is like "*x* is far from *y*"; in both cases, the claim is without sense unless construed as implicitly involving a comparison.

However, a consequence of this argument is that resemblance cannot be a relation. In the claim "*x* resembles *y*," resemblance does not function as a dyadic relation. The claim really involves four, not two entities; but neither is resemblance a tetradic relation. Our claim is not that *x*, *y*, *w*, and *z* all resemble each other; it is rather that *x* and *y* (taken as a pair) are related to *w* and *z* (taken as a pair) by the relation, not of resemblance, but of degree of resemblance. All relations, however, "have a definite, clear and unequivocal number of terms"; therefore, Butchvarov concludes, resemblance cannot be a relation.

But how does this constitute a refutation of the Resemblance Theory? Could not the proponent of that view agree that resemblance is not a relation and hold, nevertheless, that the recurrence of a quality is merely the resemblance holding between numerically different particular qualities? Butchvarov thinks not. In this context, he discusses the parallel claim "*x* is far from *y*." This claim really means, "The distance between *x* and *y* is greater than the distance between *w* and *z*." Similarly, he argues, "*x* resembles *y*," where *x* and *y* are qualities, means "The universal of least generality instantiated in *x* and *y* is of lower generality than the universal of least generality instantiated in *w* and *z*." The Identity Theory, in short, is correct in its account of the recurrence of qualities.

Despite its originality, Butchvarov's argument is somewhat less than convincing. His claim is that propositions of the form "*x* resembles *y*" are necessarily incomplete; but his only argument for this claim consists of a few remarks about the wide variation of degrees possible in resemblance. The same remarks would seem to apply to notions far less problematic than resemblance. Love, for example, is a relation that "admits of extremely wide variation of degrees"; but it does not follow that claims of the form "*x* loves *y*" are incomplete. Butchvarov does speak of a distinction between necessarily comparative relations and relations where comparison is always possible but never necessary, but his sketchy remarks about variation in degree hardly serve to make clear the required distinction.

---

[14] *Resemblance and Identity* (Bloomington, 1966).

But even if we grant that propositions of the form "*x* resembles *y*" are incomplete in the way Butchvarov suggests, it is not obvious that the introduction of another case of resemblance will enable us to complete the sense of these propositions. If "*x* resembles *y*" is incomplete, so also is "*w* resembles *z*"; but, then, one might argue, we would only add to the incompleteness of "*x* resembles *y*" by introducing the sort of comparison Butchvarov says we must introduce.

Probably the most celebrated version of realism to appear in recent years is found in W. V. Quine's *Word and Object*.[15] Quine argues that unless we admit classes into our ontology, we impoverish mathematics to the point of leaving only "some trivial portions such as elementary arithmetic." More advanced mathematics requires quantification over classes; and by Quine's criterion of ontological commitment ("To be is to be the value of a bound variable") this commits anyone who does not wish to "accommodate his natural sciences unaided by mathematics" to the existence of classes.[16]

In the light of Quine's earlier aspirations for a language compatible with the most austere nominalism, his admission of classes does appear radical. Nevertheless, too much can be made of this "conversion." Classes are the only abstract entities Quine is willing to recognize; and surely among abstract objects, classes are the least obnoxious to the "lover of desert landscapes." Unlike the abstracta of traditional realists (attributes, propositions, etc.), classes have a "crystal-clear identity concept"; classes are the same if and only if they have the same members. The account of *Word and Object* can be called realistic, but a less exciting version of realism is hardly imaginable.

B. Nominalism

During the past decade or so, few philosophers have been willing to defend a thorough-going nominalism. The most severe critics of metaphysical realism have, more often than not, been equally severe in their criticism of nominalism. Thus, Renford Bambrough, in his well-known paper "Universals and Family Resemblances," utilizes Wittgenstein's remarks about family resemblances to undermine the claims of both the realist and the nominalist.[17] According to Bambrough, Wittgenstein showed (1) that the realist is wrong to suppose that an invariant set of properties is associated with every general term and (2) that the nominalist is wrong in claiming that things called by a general term have in common only that general term. Bambrough thinks, however, that Wittgenstein meant to do more than merely criticize traditional views. Supposedly, Wittgenstein provided a positive solution to the problem of universals. As Bambrough expresses it, Wittgenstein's solution consists in his insight that things called by a general term '*F*' have in common the fact that they are all *F*'s.

Manley Thompson's discussion (in "Abstract Entities and Universals") goes a step further.[18] Thompson rejects realism and nominalism not because they are the wrong solutions to a genuine philosophical problem, but because they both make a show of resolving something that cannot be a problem at all. As Thompson interprets it, the problem of universals concerns what he calls

---

[15] *Word and Object* (Cambridge, Mass., 1960). For a discussion of Quine's views here, see C. S. Chihara, "Our Ontological Commitment to Universals," *Nous*, vol. 2 (1968), pp. 25–46.

[16] The criterion was originally presented in the classic paper, "On What There Is," *From A Logical Point of View* (Cambridge, Mass., 1953). Most recently, Quine defends the criterion (in a modified version) in "Existence and Quantification" (originally published in Joseph Margolis, *Fact and Existence* [Toronto, 1966] and later reprinted as ch. IV of *Ontological Relativity* [New York, 1969]). As he points out in that paper, his criterion of ontological commitment presupposes a referential or objectual, as opposed to a substitutional, interpretation of quantification. Very roughly, the referentialist wants to construe objects in the world as the values of bound variables. He wants to say, for example, that an existential quantification is true if and only if there is at least one object satisfying the predicates embedded in the quantification. The substitutionalist, on the other hand, takes linguistic expressions (names of objects) as the values of bound variables; he says that an existential quantification is true if and only if there is a name which, when substituted for the variable bound by the existential quantifier, makes the open sentence following that quantifier true. Quine claims that "where substitutional quantification serves, ontology lacks point;" he has, however, reasons which are independent of the question of ontological commitment for rejecting a substitutionalist interpretation of quantification.

[17] *Proceedings of the Aristotelian Society*, vol. 61 (1960–1961), pp. 207–222. A great deal of material has appeared on this topic in recent years. Some papers include K. Campbell, "Family Resemblance Predicates," *American Philosophical Quarterly*, vol. 2 (1965), pp. 238–244; Haig Khatchadourian, "Common Names and Family Resemblances," *Philosophy and Phenomenological Research*, vol. 18 (1957–1958), pp. 341–358; L. Pompa, "Family Resemblances," *The Philosophical Quarterly*, vol. 17 (1967), pp. 63–69; Robert Richman, "Something Common," *The Journal of Philosophy*, vol. 59 (1962), pp. 821–830.

[18] *Mind*, vol. 76 (1967), pp. 365–381.

Platonic entities, patterns or recipes accounting for the sameness among things.[19] The fact is that we make generalizations; we say that all the things of a kind are, in relevant respects, the same. These generalizations, however, go beyond any evidential base experience can provide. Thus, the question arises: how can we know that the generalizations are true? According to Thompson, the realist answers the question by claiming that we have knowledge of Platonic entities, after the image of which all the things of a kind are patterned; the nominalist, on the other hand, rejects Platonic entities and seeks to account for our ability to make these generalizations in other ways.

Thus, the realist and the nominalist agree that while there are certain things we can know on the basis of observation, the relevant generalizations are not among them. Thompson finds this view incoherent; he argues that all knowledge, whether observational or not, presupposes the generalizations at issue. The attempt, then, to justify our knowledge of such generalizations is really the attempt to justify knowledge in general. But, according to Thompson, that is a hopeless task; it involves seeking "evidence for what we presuppose when we offer evidence."

Almost the reverse of the Bambrough-Thompson line is defended in one of Wolterstorff's early papers, "Qualities."[20] Wolterstorff argues that realism and nominalism are both complete, internally consistent positions and that, consequently, there is no basis for choosing between them. The paper is interesting because, in attempting to establish his thesis, Wolterstorff provides one of the few defenses of nominalism to be found in recent philosophical literature. Like Shwayder, Wolterstorff construes the analysis of abstract singular terms as central to the realism-nominalism debate; however, unlike Shwayder, he thinks that the nominalist can offer a cogent account of the referring uses of these expressions: the nominalist can construe abstract singular terms as referring to what Wolterstorff calls quality-classes. A quality-class is a class of particulars such that (1) every member of the class is similar to every other and (2) no non-member is similar to all of the members.

Wolterstorff notes, however, that this definition will prove unsatisfactory if we limit membership in quality-classes to physical objects. To take one example, it could happen that all green things are round and vice versa; but if only physical objects are to be included in quality-classes, the definition would force us to construe greenness and roundness as one and the same thing. According to Wolterstorff, such difficulties disappear when we recognize that the category of particulars includes not only physical objects, but also what he calls aspects or abstract particulars. Roughly, Wolterstorff's aspects are Aristotle's first accidents, things like the roundness of this particular ball and the redness of that particular shirt. If we admit entities like these into our quality-classes, the original definition works. For every quality of the realist, we have a distinct quality-class which we can take to be the possible referent of some abstract singular term.

One might want to challenge Wolterstorff's claim that he is presenting us with a form of nominalism. We generally think of the nominalist as the philosopher whose ontology does not incorporate abstract entities of any sort. Such theoretical parsimony, however, is surely lacking in Wolterstorff's account. Admittedly, his nominalist eschews common qualities, but that's about all. He is committed at once to classes and abstract particulars. Quine, in admitting only classes, construes himself as having abandoned the nominalist cause; compared with Quine, Wolterstorff's nominalist is a full-blown Platonist.

But, labels aside, there is a genuine difficulty in the view Wolterstorff defends. He wants to claim that abstract particulars enter into quality-classes with the physical objects that possess them; and on his account, things belonging to a single quality-class are similar. But does it even make sense to say, for example, that the redness of my shirt and the shirt itself are similar? The difficulty lies in specifying how they are similar; there would appear to be nothing we could say about the two things that would justify the claim that they are similar. Possibly, Wolterstorff would want to say that they are both red; but surely this is wrong; for *pace* Plato, the redness of my shirt is not itself red.

The only fully developed nominalistic account to appear in recent years is found in the writings of

---

[19] Thompson distinguishes between Platonic entities and abstract entities (things like numbers, colors, etc.). In an earlier paper, "Abstract Entities," *The Philosophical Review*, vol. 69 (1960), pp. 331–354, he argues that as regards abstract entities, to be is to be an element in a system. In the present paper, he alludes to this criterion and construes it as establishing that abstract entities present no real problem for the nominalist. Commitment to the existence of an abstract entity need not involve more than commitment to the existence of a certain system (e.g., elementary arithmetic or color theory).

[20] *The Philosophical Review*, vol. 69 (1960), pp. 183–200.

Wilfrid Sellars. In "Grammar and Existence: A Preface to Ontology," Sellars argues that logic and mathematics do not commit us to the existence of abstract entities.[21] The fact that both involve quantification over variables of higher levels than individual variables is not decisive. According to Sellars, Quine's dictum, "To be is to be the value of a bound variable," rests on a misinterpretation of the logic of quantification. The existential quantifier is existence-involving only in a narrow range of cases; and, Sellars argues, the relevant quantifications in logic and mathematics are not among these.[22]

In the same paper, Sellars argues that ordinary talk about meaning is ontologically neutral. The Platonist claims that talk about meaning commits us to the existence of abstract entities; but according to Sellars, this is to misconstrue what we do when we state the meaning of a word. Talk about meaning generally conforms to the translation rubric, expressed by the schema

'————' (in L) means . . .

Thus, " 'Dreieckig' (in German) means triangular." Using the translation rubric, we do not refer to abstract entities of any kind; we merely exhibit the use of a word (here, "Dreieckig") by using, not mentioning, another word (here, "triangular") with presumably the same use.[23]

Sellars would agree with philosophers like Shwayder and Wolterstorff that the major obstacle to nominalism is presented by abstract singular terms. He feels, however, that in many contexts the use of abstract singular terms is non-problematic. Frequently we use abstract terms merely by way of abbreviation. In such contexts, what we say by using abstract singular terms could just as well be said by using only concrete forms. Thus, "triangularity" in "Triangularity implies having three sides" presents no real problem for the nominalist; for, according to Sellars, the sentence in which it appears is simply another way of saying that if anything is a triangle, it has three sides.

Sometimes abstract singular terms enter into talk about exemplification. Thus, we might say that some object exemplifies triangularity. Sellars is one of the few nominalists (possibly the only one) to grant the truth of such remarks; but as he argues in "Naming and Saying," talk about exemplification is merely another way of talking about the truth of sentences. Thus, "x exemplifies triangularity" is analyzed as "Triangularity is true of x"; and this, in turn, becomes "That x is triangular is true."[24]

In some of their uses, however, abstract singular terms cannot be disposed of so easily. We say, for example, that triangularity is a quality and that mankind is a class. In such cases, we have what might be called essential occurrences of abstract singular terms. In an essential occurrence, an abstract singular term functions as subject; but it seems impossible to construe the expression as a stand-in for one or more concrete terms. What we say about the thing designated by the subject is something that can be correctly said only about an object that is, in a quite straightforward sense, an abstract entity.

In "Abstract Entities" Sellars attempts to deal with essential occurrences of abstract singulars.[25] He wants to claim that, in essential occurrences, abstract singular terms signify what he calls linguistic types. A linguistic type is a word "construed as a kind or sort which is capable of realization or embodiment in different linguistic materials." Qua kind or sort, it exhibits the linguistic role that expressions from historically different languages can have in common. Thus, "dreieckig" and "triangular" are different realizations of one and the same linguistic type. Sellars uses a dotting device for referring to linguistic types. "Dreieckig" and "triangular," he tells us, are both ·triangular·s; they both realize the linguistic type, the ·triangular·. It is this linguistic type that the abstract singular term "triangularity" signifies, so that in an essential occurrence, "triangularity" is synonymous with "the ·triangular·."

---

[21] *Mind*, vol. 69 (1960), pp. 499–533; also found in *Science, Perception, and Reality*, *op. cit.*, ch. VIII.

[22] Very generally, Sellars' rejection of Quine's criterion of ontological commitment has its source in Sellars' unwillingness to accept a referential interpretation of quantification. It is not altogether clear whether Sellars is a substitutionalist, but he does think it a mistake to read the existential quantifier in the traditional way as "There exists. . . ." He suggests instead that we read the quantifier as "Some . . ."; and he contends that this is the proper reading regardless of the style of variable bound by the quantifier. Nevertheless, he wants to say that the existential quantifier does play the role of ascribing existence in those cases where the variables it binds take as their values singular terms. Sellars' inclination to call linguistic expression the values of variables leads one to believe that he is an out-and-out substitutionalist. My doubts about the label are due to some negative remarks I have heard Sellars make about the appropriateness of the term.

[23] For further discussion of this and other topics central to the realism-nominalism debate, see Sellars' *Science and Metaphysics* (London, 1968). Although this book provides a synthesis of Sellars' views on a wide variety of topics, it tends to be very difficult reading and is best approached only after one has an acquaintance with some of Sellars' earlier papers.

[24] The proposed translation might be contested on the grounds that it makes reference to a proposition and is not, therefore, completely nominalistic. Sellars would argue, I think, that propositions can be handled by the method of linguistic types, described in the following paragraphs.

[25] *The Review of Metaphysics*, vol. 16 (1963), pp. 627–671; also reprinted as ch. IX of Sellars' *Philosophical Perspectives* (Illinois, 1967).

One might object that Sellars has merely replaced one kind of abstract object for another; for if linguistic types have different realizations, different embodiments, then surely they are to be reckoned universals. Sellars' contention, however, is that "the ·triangular·" does not signify a universal; it is rather what he calls a distributive singular term. It functions in the way "the lion" functions in "The lion is tawny." Here, we are not referring to the universal, the lion; for while "The lion is tawny" is true, "The universal, the lion, is tawny," if intelligible at all, is false.

But what is a distributive singular term? According to Sellars, to use an expression of the form "the *K*" as a distributive singular term is not to refer to something over and above the things that are *K*'s; it is rather to make a logically necessary statement about all *K*'s. Thus, all sentences of the form "The *K* is *F*" (where "the *K*" functions as a distributive singular term) are reducible to sentences of the form "All *K*'s are necessarily *F*."[26] Thus, to speak of the ·triangular· is not to speak of something over and above expressions that are ·triangular·s; it is simply to make a necessary claim about all those expressions in historically different languages which play the role that "triangular" plays in our language.

This analysis, however, disposes of only half of our original difficulty. We were concerned with sentences like "Triangularity is a quality" and "Mankind is a class." The terms "triangularity" and "mankind" are, on Sellars account, ways of referring to ·triangular·s and ·man·s; but what of the expressions "quality" and "class" as they appear in these sentences? Here Sellars appropriates Carnap's talk of quasi-syntactical sentences. Our examples are sentences in the material mode; but actually, they are remarks about the syntax of ·triangular·s and ·man·s; this is brought out, Sellars contends, when we translate them into the formal mode. "Triangularity is a quality" becomes "The ·triangular· is an adjective" and this, in turn, becomes "All ·triangular·s are necessarily adjectives." In similar fashion, "Mankind is a class" becomes "The ·man· is a common noun" and, finally, "All ·man·s are necessarily common nouns."

Unfortunately, Sellars is never clear about the force of his nominalism. Sometimes it appears that his analyses are meant to capture existing synonymies; but in other contexts Sellars seems to find the notion of strict translation too confining and speaks instead of reconstructing language along nominalist lines. Clearly, these are two quite different things; and unless we know which of these aims is, in the final analysis, operative, we can hardly assess the particular moves in Sellars' account.

Aside from this difficulty, Sellars' notion of a linguistic type is troublesome. For one thing, Sellars appears to have Quine's remarks about the indeterminacy of translation to contend with; for certainly, one consequence of Quine's arguments is that it is impossible to identify linguistic roles across languages in the way Sellars claims we can.[27] More important, however, is the fact that when we eliminate a linguistic type in favor, for example, of materially different ·triangular·s, we seem to be left with a universal, the linguistic role that all ·triangular·s have in common. It will not do, of course, to suggest that the linguistic role that ·triangular·s share is itself susceptible to elimination via distributive singular terms; for this presupposes the introduction of what is also, *prima facie*, a universal, the linguistic role underlying the linguistic type, the ·linguistic role·.

Sellars, I suspect, would argue that linguistic roles need not be eliminated in this way. He would doubtless claim that we can explain all that is involved in a particular linguistic role by specifying what he calls the prescriptive criteria for the use of the word in question, by specifying, that is, all the moves that can be made in the language-game with that word. The difficulty here is that we have no guarantee that a specification of prescriptive criteria will not involve new universals and that in eliminating these, still others will not emerge.

II. PARTICULARS, SUBSTANCES, AND INDIVIDUATION

In this section I want to examine the cluster of problems surrounding the notion of a particular.[28] I shall first deal with the appeal to bare particulars found in the writings of Gustav Bergmann and his followers; then, I shall examine the work of philosophers who want to construe particulars as complexes of properties or qualities. In the third section I shall discuss the essentialist interpretations of particulars found in writers like Plantinga, Geach, and Wiggins; and finally, in the fourth section, I shall deal with the account of particulars provided by Strawson in *Individuals*.

---

[26] For a criticism of Sellars' reduction-schema for distributive singular terms, see Wolterstorff's "On the Nature of Universals," discussed earlier in this survey.

[27] Quine's views on this topic are expressed in ch. II of *Word and Object*, *op. cit.*, and ch. I–III of *Ontological Relativity*, *op. cit.*

A. Bergmann on Individuation

In all of his writings, Bergmann is critical of substance ontologies. According to Bergmann a substance ontologist is committed to the view that objects can change while remaining numerically identical; and identity through change, Bergmann argues, is *prima facie* contradictory.[29] The substance ontologist is committed to the view that a substance, *a*, both has and has not a certain character, *F*. To eliminate this contradiction, the substance ontologist must introduce temporal moments and say that *a* has *F* at *t* and *a* does not have *F* at $t+n$. Unfortunately, Bergmann claims, the introduction of temporal moments violates the Principle of Acquaintance. According to that principle an ontology must meet the demands of a strict empiricism; it may posit only such entities as are objects of possible acquaintance; but surely, Bergmann contends, we are never acquainted with non-relational temporal moments.

Bergmann concludes that a sound ontology must construe the items out of which our common-sense world is constituted as changeless. In his earlier writings, Bergmann found in this claim support for phenomenalism, the view that our world is a construction of changeless sense-data. He has since rejected phenomenalism; but he still maintains that whatever we take to be the building-blocks of reality, we must construe them as changeless.[30]

When we reach the level of sense-data or their non-phenomenological counterparts, we have not yet, in Bergmann's view, reached the level of ontological simples. As our discussion of Bergmann on universals indicates, the level of simples incorporates talk about individuals and their characters. The relevant individuals, of course, are what Bergmann calls bare particulars. In his most recent book, *Realism*, Bergmann again presents the case for bare particulars; and here, as in his earlier writings, the notion of a bare particular is tied to considerations about numerical diversity.[31] It is logically possible, Bergmann contends, that two objects have all their non-relational characters in common. This implies, he argues, that objects are not constituted exclusively by their characters. We cannot account for the individuation of objects by reference to their relational characters. Admittedly, qualitatively identical objects will differ relationally; but according to Bergmann, relations presuppose, rather than explain, numerical diversity. To account for the possibility of objects agreeing in all their non-relational characters, we must grant that each object is constituted not only by a set of characters, but also by some non-repeatable entity (a bare particular) to which we can attribute the identity of that object.

One might wonder whether bare particulars, any more than pure temporal moments, are compatible with the Principle of Acquaintance; for surely we are never acquainted with entities lacking all characters. Bergmann agrees that we never perceive bare particulars by themselves; we only meet with particulars as exemplifying characters; and this is connected with what he calls the Principle of Exemplification, the claim that particulars are never uncharacterized, nor characters unexemplified. Nevertheless, Bergmann contends that we can point to something in perception we can call bare particulars. Bergmann's view is clearly expressed in Edwin Allaire's "Bare Particulars."[32] According to Allaire, a bare particular just *is* the thing responsible for the individuation of objects alike in all their non-relational characters; it just *is* the carrier of numerical difference; but then, in being acquainted with numerically different objects exemplifying the same non-relational characters, Allaire argues, we are *eo ipso* acquainted with bare particulars.

---

[28] For a general account of the problems considered in this section, see Michael Loux, "Particulars and Their Individuation," in *Universals and Particulars*, *op. cit.* See also Douglas Long, "Particulars and Their Qualities," *The Philosophical Quarterly*, vol. 18 (1968), pp. 193–206.

[29] See ch. XI and XIV of *Meaning and Existence*, *op. cit.*, ch. V of *Logic and Reality*, *op. cit.*, and ch. VI of *Realism*, *op. cit.* For discussions of Bergmann's views here, see Reinhardt Grossman, "Particulars and Time," Essays in Ontology and John Kearns, "Substance and Time," *The Journal of Philosophy*, vol. 67 (1970), pp. 277–289.

[30] For Bergmann's account of his rejection of phenomenalism, see "Realistic Postscript" in *Logic and Reality*, *op. cit.*

[31] See especially ch. IV of *Realism*, *op. cit.*

[32] *Philosophical Studies*, vol. 14 (1963), pp. 1–8. This paper spawned an interesting controversy. V. C. Chappel ("Particulars Re-Clothed," *ibid.*, vol. 15 [1964], pp. 60–64), criticized Allaire's account and suggested that bare particulars are unnecessary as individuators since space and time suffice to individuate qualitatively indistinguishable objects. Allaire responded ("Another

But there is a difficulty here. As the Principle of Exemplification makes clear, bare particulars are not merely the carriers of numerical diversity; they are also the items which, in the final analysis, exemplify characters. Considered in this light, Bergmann's notion of bare particulars seems less compatible with his empiricist criterion of ontology. The proponent of the Principle of Acquaintance may agree with Bergmann and Allaire that we are acquainted with bare particulars as carriers of numerical identity; but he will likely challenge the claim that we are acquainted with bare particulars qua possessors of characters; and here he will merely be echoing Bergmann's contention that perception has as its object a *quale*, never a *quid*. What is needed is an argument to support the claim that the items responsible for the individuation of objects are also the items exemplifying the characters we associate with those objects; but as far as I can tell, Bergmann never presents such an argument.

The difficulty goes beyond the Principle of Acquaintance. It is not simply that experience does not reveal bare substrata to be the carriers of characters. The further question is whether we can even make sense of the notion of an unqualified possessor of qualities. In "Particulars," Wilfrid Sellars argues that Bergmann's view is contradictory,[33] Sellars invites us to translate into the symbolism of *Principia Mathematica* the claim that whatever exemplifies a character is itself bare. The result, he contends, is an out-and-out contradiction:

$$(\forall x)\ \{\ [(\exists\psi)\ (\psi x)] \supset [-(\exists\phi)\ (\phi x)]\ \}\ .$$

Similarly, Bergmann's attack on substance ontologies is problematic. Bergmann's remarks on change and time are deeply rooted in his Ideal Language method. Why, for example, must the substance ontologist construe time as a series of non-relational moments? Bergmann's answer is that this is the only account of time which, in conformity with the syntax of the Ideal Language, enables the substance ontologist to eliminate the contradiction inherent in change; and it is the syntax of the Ideal Language which proves central in the argument. Assuming that Bergmann is right on this point, the philosopher operating from outside Bergmann's system is likely to be impressed less with the inadequacies of substance ontologies than with the inadequacies of Bergmann's Ideal Language. If the syntax of a language is such that it permits us to say what we mean about change only if we assume an untenable view about time, then, we want to say, that language is hardly one that is "ideal" for all philosophical purposes.

B. *"Clusters" and "Bundles"*

Probably the most persistent critic of Bergmann's account of individuation is Herbert Hochberg. According to Hochberg, the bare particular analysis cannot, in any way, be reconciled with the Principle of Acquaintance.[34] On the contrary, the strict empiricist must claim that objects are merely complexes of universals, "bundles" or "clusters" of characters. Hochberg's main contention is that a view of this sort can handle difficulties about individuation at least as well as the bare particular analysis. In "Things and Qualities," Hochberg argues that complexes of universals are individuated of and by themselves.[35] He claims that bare particulars are posited merely to provide the ontologist with entities whose numerical diversity requires no further explanation. Supposedly, they do not differ "in anything; they simply differ." But why not hold, Hochberg asks, that complexes of characters are the sorts of things that do not differ "in anything," but "simply differ"?

In "Things and Descriptions" Hochberg presents a rather different account.[36] There he contends that relations individuate. Numerical difference among objects alike in all their non-relational features has its source in the various

---

Look at Bare Particulars," *ibid.*, vol. 16 (1965), pp. 16–21) with the claim that differences in spatio-temporal location themselves presuppose numerical diversity and, therefore, bare particulars, a claim which J. W. Meiland criticized in "Do Relatives Individuate?" *ibid.*, vol. 17 (1966), pp. 65–69. The controversy closed with Allaire's brief response, "Relations and the Problem of Individuation," *ibid.*, vol. 19 (1968), pp. 61–63.

[33] *Philosophy and Phenomenological Research*, vol. 13 (1952), pp. 184–199; reprinted as ch. IX of *Science, Perception, and Reality*, *op. cit.* A similar argument is found in Jonathan Bennett, "Substance, Reality, and Primary Qualities," *American Philosophical Quarterly*, vol. 2 (1965), pp. 1–17.

[34] See Herbert Hochberg's "Ontology and Acquaintance," *Philosophical Studies*, vol. 17 (1966), pp. 49–55.

[35] Herbert Hochberg, "Things and Qualities" appeared in W. H. Capitan and D. D. Merrill, *Metaphysics and Explanation* (Pittsburgh, 1964).

[36] "Things and Descriptions," *American Philosophical Quarterly*, vol. 3 (1966), pp. 39–47.

relations obtaining among such objects. Hochberg agrees with Bergmann that relations presuppose numerical diversity; but he claims that the dependence may operate in both directions. Admittedly, relations can hold only between numerically distinct objects; but it may also be that numerical diversity, in turn, presupposes the existence of one or more relations.

Although critical of the appeal to bare particulars, Hochberg finds certain features of Bergmann's ontology indispensable. Thus, he denies that particulars are mere heaps of universals. One must postulate, he argues, a structural tie or connection (analogous to Bergmann's nexus of exemplification) among the various characters constituting an object. This tie accounts for the unity or connexity we find in objects and enables us to explain why some logically consistent sets of characters actually constitute objects, while others do not.[37] Further, Hochberg's objects, like Bergmann's, are changeless. Change involves an alteration in the characters associated with an object; but since objects just are these characters, change necessarily yields a numerically different object.

An interesting feature of Hochberg's account is his analysis of reference. It is frequently claimed that a "bundle" theorist is forced to construe most of our ordinary talk about objects as analytic. If an object just is a complex of characters, then to know which object is being referred to, a person must know in advance all of the characters that can be truly predicated of that object. But if this is true, no character could ever be truly, yet non-tautologously, applied to an object. Hochberg denies that this is a real consequence of his view. In "Things and Qualities" he focuses on proper names. Proper names, he contends, are not mere labels for lists of characters. Indeed, they have no meaning at all, only a reference; nor is the referential connection between name and object effected by means of an object's characters; the connection is immediate. Thus, a true subject-predicate proposition in which a proper name functions as subject is not analytic; a person can understand the proposition and know which object it is about, without knowing in advance that it is true.

In "Things and Descriptions" Hochberg deals with propositions in which we refer to objects by means of definite descriptions. In a proposition of this sort, we pick out an object as the thing which has *F*, *G*, *H*, etc.; we focus on such characters as are jointly sufficient for uniquely identifying the object. Since relations individuate, the relevant characters will be only some of the many characters exemplified by an object. Thus, we can go on and ascribe to an object so described other characters; and the resulting propositions will be synthetic. But, Hochberg argues, propositions in which we identify an object by a definite description and then ascribe to it one or more of the characters introduced in that description will also be synthetic. Certainly, propositions of this sort will be redundant; but since they incorporate a definite description in the role of grammatical subject, they are really only disguised existential propositions and, therefore, cannot be analytic.

Like Hochberg, Neil Wilson ("Substances without Substrata") is critical of the notion of a bare particular.[38] As Wilson interprets the doctrine, the proponent of substrata is committed to the view that objects are what they are regardless of the properties they possess. In refutation of this claim, Wilson asks us to imagine a world like ours in all respects except that all of the properties of Marc Antony belong to Julius Caesar and vice versa. If the doctrine of substrata were correct, the resulting world would be different from ours; but, according to Wilson, the imagined world would, in no way, differ from our own world; it would not represent a new possibility—the substrata theory must be wrong.

Wilson is more sympathetic with the view that objects are merely "bundles" of properties. The "bundle" theorist, he contends, is right in construing the identity of objects as dependent upon the properties they possess; and in this connection, Wilson defends Leibniz' Identity of Indiscernibles, the principle that no two objects can exhibit the same set of properties. But, in the final analysis, Wilson finds the bundle analysis unsatisfactory. For one thing, the bundle theorist is committed to the view that objects cannot remain numerically identical through change; and Wilson, unlike Hochberg, thinks that this is too great a price to pay. Further, the "bundle" theorist wants to construe properties as concrete, as independent existents having an identity all their own; but according to Wilson, neither objects nor their qualities are concrete in this sense. It is only facts that are what they are independently of other things. But is each individual fact concrete? Wilson thinks not; atomism, he contends, must give way to some form of Absolute Idealism. In the end, only a very large number of facts, possibly the totality of facts, constitutes a concrete existent.

---

[37] The question of the relations holding among the qualities of an object is considered by Haig Khatchadourian in "Objects and Qualities," *American Philosophical Quarterly*, vol. 6 (1969), pp. 103–115. He contends that it is incorrect to speak of relations here. The consequences for the "bundle" theorist should be obvious.

[38] *The Review of Metaphysics*, vol. 12 (1959), pp. 521–539. For a discussion of Wilson's "test-case," see A. N. Prior, "Identifiable Individuals," *The Review of Metaphysics*, vol. 13 (1960), pp. 684–696.

Wilson's position, then, is not a version of the "bundle" theory; but, apart from the account presented by Hochberg, it is about the closest approximation to the "bundle" theory to be found in recent literature. Historically (although not logically), the "bundle" theory has been associated with Phenomenalism; and the decline, in recent years, of the latter view has left the former position with almost no proponents. Indeed, even A. J. Ayer, who earlier construed the bundle theory as obviously correct ("The Identity of Indiscernibles"), has become more cautious.[39] In "Names and Descriptions," Ayer argues that our language, as it stands, implies an ontology of Lockean substrata.[40] He seems to think that it is possible to construct a language free of this commitment; but even this, he argues, would not warrant adherence to the view that objects are simply "bundles" of properties. The "bundle" analysis commits one to the Identity of Indiscernibles; and Ayer is now willing to question (as he earlier was not) the truth of that principle.

Technical difficulties aside, it is not surprising that philosophers have been reluctant to adopt some version of the "bundle" theory. On that theory, ordinary objects are constructions out of qualities; the difficulty with this claim is that many of the qualities involved in the proposed construction would appear to be object-laden themselves. G. E. M. Anscombe argues persuasively for this point in "Substance."[41] She grants that some of the qualities associated with a substance are not substance-involving; i.e., they do not presuppose the concept of substance. According to Miss Anscombe, most of Locke's secondary qualities are of this sort. Thus, there could be a patch of color which did not constitute the surface of any physical object. However, secondary qualities are not, by themselves, sufficient for the construction of our substance-concepts. We need other qualities; and these, Miss Anscombe argues, are not logically independent of the notion of substance. Thus, qualities like malleability and *melts at* 44°*C* are substance-involving. These qualities can be instantiated only as the qualities of this or that physical substance. But surely, she concludes, it would be absurd to construe qualities like these as items out of which our substance-concepts are constructed; they already presuppose those concepts.

C. Essentialism and Identity

Some philosophers have wanted to distinguish between properties which are essential and properties which are only accidental to a particular. Very generally, the distinction is as follows: while an object can change with respect to its accidental properties, its essential properties are inseparable from the object; to lose one of them is to become numerically a different object. Essentialists (as proponents of this distinction are called) have, almost without exception, been critical of the two sorts of accounts of particulars we have so far examined. Against proponents of bare particulars, essentialists have argued that bare substrata can never function as logical subjects; the ultimate subjects of predication, they argue, are particulars essentially characterized. In opposition to the "bundle" theorist's radical denial of identity through change, essentialists have argued that where only accidental properties are involved, particulars can and do change while remaining numerically the same.

Essentialists, however, have disagreed among themselves about the nature of essential properties. Some (Leibniz, for example and probably Scotus) have spoken of individual essences, essences peculiar to particular objects; whereas others (most notably, philosophers in the Aristotelian tradition) have argued that essences are always general, common to all the members of a kind or sort.

In recent philosophy, the Leibnizian tradition is represented by Alvin Plantinga. In "World and Essence," he defends a notion of individual essence that is equally as rich as Leibniz' notion of haecceity or complete individual concept.[42] Before examining that paper, however, I want to look at the defense of essentialism presented in *God and Other Minds*.[43] Here, Plantinga does not defend any particular version of essentialism; he is concerned with the more general task of clarifying the distinction between essential and accidental properties. His claim is that we can explicate the notion of a necessary or essential property (*de re* necessity) in terms of the less problematic notion of necessity at work in talk about propositions (*de dicto* necessity). We can say that a property, P, is essential to an individual, x, if and only if x has P and the proposition x *lacks* P is necessarily false.[44]

---

[39] Included in Ayer's *Philosophical Essays* (London, 1953).

[40] Found in Ayer's *The Concept of a Person* (London, 1964).

[41] *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVIII (1964), pp. 69–78. For a discussion of Miss Anscombe's paper see Stephan Körner, "Substance," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVIII (1964). pp. 79–90.

[42] *The Philosophical Review*, vol. 79 (1970), pp. 461–492.

Plantinga argues, however, that this characterization is adequate only if we place certain restrictions on the substituend sets of the variables 'x' and 'P'. In its unrestricted form, the characterization forces us to construe properties as both accidental and essential to one and the same object.[45] Thus, "Seven lacks the property of being prime" is necessarily false; *being prime* is a property essential to seven. Now, suppose that seven is Quine's favorite number. Applying the criterion, *being prime* turns out to be accidental to seven; for while "Quine's favorite number lacks the property of being prime" is false, it is not necessarily false. The difficulty here lies in our ability to refer to the number seven in a variety of ways. The same difficulty can arise in the case of the property *being prime*. Thus, suppose *being prime* is the property that Quine is now thinking of. Again, the criterion forces us to construe *being prime* as accidental to seven; for "Seven lacks the property Quine is now thinking of" does not express a necessary falsehood.

According to Plantinga, the renegade labels are all definite descriptions; if we limit the substituend sets of the variables 'x' and 'P' appearing in the original characterization to proper names, the characterization works.[46] But, then, taking the criterion, as restricted in this way, we can accommodate cases where properties and/or the objects possessing them are identified by definite descriptions. We can say: a property, P, is essential to an object, x, if and only if (1) x has P and the proposition x lacks P (where the substituend sets of 'x' and 'P' contain only proper names) is necessarily false or (2) x has P and there is an object, y, and a property, P', such that x is identical with y and P is identical with P' and P' is an essential property of y in the sense of (1).

In "World and Essence," Plantinga's approach is quite different. Appropriating the modal logician's talk about possible worlds, he claims that a property is essential to an object just in case that object has the property in every possible world in which it exists. Some properties, Plantinga claims, are trivially essential; thus, every object in every possible world has the property of *being colored if red*. Other properties are essential to some objects, while accidental to others; *being Socrates or Greek*, for example, is essential to Socrates, but accidental to Plato and Aristotle. Still other properties—*being Socrates or Plato*, for example—are essential to all the objects possessing them.

In addition to these kinds of essential properties, there is for each object, Plantinga argues, an haecceity or individual essence, a property which (1) is essential to the object in question, (2) is possessed only by that object, and (3) entails all of the other properties essential to that object. Thus, Socrates has *Socrates-identity*. (1) This property is essential to Socrates—there is no possible world in which Socrates both exists and lacks this property; (2) no object distinct from Socrates has, in any possible world, *Socrates-identity*; (3) the property entails all of Socrates' other essential properties—there is no possible world in which Socrates has *Socrates-identity* but lacks one or more of his other essential properties.

Plantinga, however, wants to go much further than this. He wants to say, with Leibniz, that an Omniscient Being could, by reflecting on Socrates' essence, "read off" all of Socrates' properties, both accidental and essential. Suppose, for example, that Socrates actually was snub-nosed in the real world. *Being snub-nosed* will doubtless be accidental to

---

[43] *God and Other Minds* (Ithaca, 1967).

[44] An expanded version of this account is found in Plantinga's "*De Re et De Dicto*," *Nous*, vol. 3 (1969), pp. 235–258.

[45] This difficulty is doubtless familiar to most readers. It is raised by Quine in a variety of places. See, for example, *Word and Object*, *op. cit.*, Sect. XLI. See Hugh Chandler, "Essence and Accident," *Analysis*, vol. 26 (1966), pp. 185–188 for a rather different treatment of the difficulty.

[46] In the interests of space, I oversimplify here. Plantinga does not hold that properties have proper names in the way persons, for example, do. There are, however, labels for properties (e.g., "courage") that are neither definite descriptions nor equivalent to such. It is these labels that are to be included in the relevant substituted sets. Further, Plantinga realizes that many objects lack proper names and many properties, uniquely referring labels of a non-descriptive sort. In *God and Other Minds*, *op. cit.*, he does not make much of this point; but in "*De Re et De Dicto*," *op. cit.*, his characterization of essential properties is explicitly formulated to handle unnamed objects and label-less properties.

Socrates; but, Plantinga argues, *being snub-nosed in the actual world* is not accidental to him. There is no world in which Socrates both exists and lacks this second-level property; and qua essential to him, the property of being snub-nosed in the actual world is entailed by Socrates' individual essence. But we could take any property of Socrates and, in the same way, derive from it a property that is essential to Socrates and, therefore, entailed by his essence. Consequently, for any property, *P*, and any world, *W*, God could infer from Socrates' essence that Socrates has or has not *P* in *W*. However, God also knows which world is the actual world, so that, Plantinga concludes, He is able to "read off" all of the properties Socrates possesses in the actual world—accidental as well as essential.[47]

In the writings of P. T. Geach, we find a version of essentialism that falls within the Aristotelian tradition of common or general essences. Very roughly, Geach's view is that nothing can function as an object of reference unless it is possible to specify in general terms (i.e., terms appropriate to other things as well) the kind of thing the object is. More often than not, Geach presents his version of essentialism in the linguistic mode. In *Reference and Generality* and *Mental Acts*, he argues that a necessary condition for the correct use of a proper name is the ability to specify some count-noun true of the object bearing the name.[48] Geach's count-nouns are expressions like "man" and "dog"; essentially, they are Strawson's sortal terms. Besides furnishing us with principles for enumerating objects, they provide a classification of objects by indicating the kinds or sorts to which objects can belong. Count-nouns, Geach wants to say, tie proper names to their referents. Proper names are meant to be used repeatedly; but if I am to use a proper name correctly on different occasions, I must know whether or not the object I am referring to is the object bearing the relevant proper name. To know this is to know that a statement of numerical identity (a statement of the form "*x* is the same as *y*") is true. But, Geach argues, the claim that two things are the same is incomplete. To fill in the claim, one must answer the question, "Same what?"; and, according to Geach, it is only by means of count-nouns that one can do this.

The talk about proper names, then, is really subsidiary. The key issue is Geach's claim that statements of the form "*x* is the same as *y*" are incomplete. But why are they incomplete? The reason, Geach argues, is that identity is relative, relative to the count-noun one specifies. It is possible, he claims, for an object, *x*, and an object, *y*, to be the same *F*, but different *G*.[49] Thus, the mayor of Loogootie is the same man as the president of the Chickasaw County Board, but a different official personage. Similarly, what I bathed in today and what I bathed in yesterday are the same river, but not the same water. But since identity is relative in this way, I can hardly know what is being asserted by the bare claim that *x* is the same as *y*. I need to know the count-noun under which *x* and *y* are being judged identical.

It should be clear that Geach's account is incompatible with traditional accounts of identity. If, as Geach argues, identity is relative to the count-noun specified, then we must, among other things, reject the principle known as Leibniz' Law. Suppose that some object, *x*, and some object, *y*, are the same *F*; their identity under '*F*' does not guarantee that whatever is true of one is true of the other. They might, despite their identity as *F*'s, be different *G*'s; and, then, it will be false, for example, to say of *y* what can be truly said of *x*—that it is the same *G* as *x*.

In his review of *Reference and Generality*, Quine roundly criticizes Geach on this point.[50] Logic, he argues, requires an absolute notion of identity, one conforming to Leibniz' Law. In a similar vein, David Wiggins (*Identity and Spatio-Temporal Continuity*) complains that if we follow Geach in rejecting Leibniz' Law, we no longer have a justification for the substitution of identicals in deductive contexts, nor do we leave ourselves with anything more than the bare notions of reflexivity, transitivity, and symmetry to define the concept of identity.[51]

---

[47] For further discussion of topics central to Plantinga's account, see Roderick Chisholm, "Identity Through Possible Worlds," *Nous*, vol. 1 (1967), pp. 1–8; Jaakko Hintikka, "Individuals, Possible Worlds, and Epistemic Logic," *Nous*, vol. 1 (1967), pp. 33–63; Leonard Linsky, "Reference, Essentialism, and Modality," *The Journal of Philosophy*, vol. 66 (1969), pp. 687–700; R. Purtill, "About Identity Through Possible Worlds," *Nous*, vol. 2 (1968), pp. 87–89.

[48] See ch. II of *Reference and Generality* (Ithaca, 1962) and ch. XVI of *Mental Acts* (London, 1957). This theme is also developed in *Three Philosophers, op. cit.*; indeed, in almost all of Geach's writings and in many of Miss Anscombe's, one finds the view operative.

[49] Again, I oversimplify here. It is not simply that an object, *x*, and an object, *y*, might be the same *F* and different *G*'s: they might be the same *F*, with one of them being a *G* and the other, not only not the same *G*, but not a *G* at all.

[50] *The Philosophical Review*, vol. 73 (1964), pp. 100–104.

Wiggins distinguishes two claims: (1) the claim that statements of the form "x is the same as y" requires a count-noun to complete their sense and (2) the claim that identity is relative to the count-noun specified. Wiggins rejects (2), but defends (1), thereby accepting a form of essentialism while preserving Leibniz' Law. His attack on (2) involves a painstaking analysis of cases where we seem to have an x and a y identical under one count-noun, but different under another. In all cases but one, Wiggins claims to show that examples purporting to conform to Geachian identity actually fail to do so. The only real counter-example to Leibniz' Law is provided by the doctrine of the Trinity. According to that doctrine, three individuals are one and the same God, but numerically different Persons. Wiggins argues, however, that since in all other contexts identity conforms to the Leibnizian model, we have no option but to reject the doctrine of the Trinity.

Nonetheless, Wiggins maintains that an essentialist interpretation of identity-statements is correct. His own defense of that interpretation is limited to identity-statements involving material bodies. According to Wiggins, to say that two material objects, x and y, are identical is to say that if one traces them through space and time, he finds them to coincide. Spatio-temporal coincidence, however, is a "purely formal notion": there can be no such thing as a general account of what it is for two material objects to coincide in space and time. Consequently, Wiggins argues, claims to the effect that a material object, x, and a material object, y, are the same require completion, completion by means of terms which make it possible for us to know what spatio-temporal coincidence is for x and y. Such terms, Wiggins argues, are count-nouns or sortal terms.[52]

In a paper entitled "Identity," Geach attempts to answer his critics.[53] His main contention is that the account of identity provided by Leibniz' Law and its converse, the Identity of Indiscernibles, is untenable. He begins by arguing that the proponent of Leibnizian identity cannot accept the traditional formulation of the two principles. As it is often expressed, the conjunction of the two principles reads as follows: an object, x, and an object, y, are numerically identical if and only if whatever is true of x is true of y and vice versa. But Geach contends that to accept this formulation and use the locution "whatever is true of..." is to land oneself in "such notorious paradoxes as Grelling's and Richard's." To avoid the semantical paradoxes, one must rather say that objects are numerically if and only if they are indiscernible with respect to their predicates. But even amended in this way, Leibnizian or absolute identity encounters difficulties. Geach asks us to consider a set of objects like the following:

(1) a
(2) a

Now, Geach asks us to suppose that there exists a language so weak in predicates that it does not enable us to distinguish between types of expressions and their tokens. In that language it will be correct to say (provided we accept the amended account of Leibnizian identity) that (1) and (2) are identical. But let us further suppose that the language is supplemented in such a way that we can make the distinction between types and their tokens. When the original language is enriched in this way, (1) and (2) turn out to be numerically different objects. Thus, the proponent of Leibnizian identity is forced to grant that merely by expanding the ideology of a language (its descriptive resources), objects previously identical become diverse; and no Leibnizian, Geach contends, would welcome this consequence.

The proponent of absolute identity might respond by saying that when, in the weaker language, we make the true claim that (1) and (2) are identical, we aren't speaking of the same objects as when we say, in the enriched language, that (1) and (2) are different. Geach, however, finds this response unsatisfactory. To make this move, he claims, is to expose oneself to "indecently burgeoning existential commitments;" it is to multiply entities to the point where "the universe . . . shows itself as a baroque Meinongian structure." According to Geach, our only option is to reject the concept of absolute identity. We must admit instead that numerical identity and discernibility with respect to predicates are compatible; but this is to grant, among other things, Geach's point that it is possible for an object, *x*, and an object, *y*, to be the same *F*, but different *G*'s.

---

[51] (Oxford, 1967). For a detailed discussion of Wiggins' book, see Sidney Shoemaker, "Wiggins on Identity," *The Philosophical Review*, vol. 79 (1970), pp. 529–544. Also see Gerald Vision, "Essentialism and the Sense of Proper Names," *American Philosophical Quarterly*, vol. 7 (1970), pp. 321–33, where both Wiggins' and Geach's views are examined.

[52] For a development and defense of Wiggins' view, see M. Loux, "Identity-Statements and Essentialism," *The New Scholasticism*, vol. 44 (1970), pp. 430–439.

[53] *The Review of Metaphysics*, vol. 21 (1967), pp. 3–12. For a critical discussion of this paper, see Fred Feldman, "Geach and Relative Identity," *Review of Metaphysics*, vol. 22 (1969), pp. 547–555. See also Geach's response to Feldman (*ibid.*, pp. 556–559) and Feldman's rejoinder (*ibid.*, pp. 560–561). Other discussions of Geach's argument in "Identity" include Jack Nelson, "Relative Identity," *Nous*, vol. 4 (1970), pp. 241–260 and John Perry, "The Same *F*," *The Philosophical Review*, vol. 79 (1970), pp. 181–200.

In "Geach and Relative Identity" Fred Feldman argues that even if sound, Geach's argument does not establish the view that identity is relative to the count-noun one specifies. The nub of Geach's argument is contained in his remarks about expanding ideology. As we expand the ideology of a language, statements of identity formerly true turn out to be false; but according to Feldman, this only commits us to the view that identity is, in some way relative; it does not commit us to any particular account of the relativity. Thus, one could handle Geach's example without rejecting Leibnizian identity. One could say that statements of identity are relative, not to a count-noun, but rather to a language as a whole. On this view, statements of the form "*x* is the same as *y*" would be, as Geach suggests, incomplete; but to fill them in, one would indicate, not the count-noun governing the identity-statement, but the language, against the ideology of which the objects are being judged identical. Construed in this way, judgments of identity would still conform to the model provided by Leibniz' Law and its converse; for objects would be numerically identical in a language if and only if they were indiscernible with respect to the ideology of the language.

It is interesting how the relativism Feldman points to parallels the account of identity found in Quine's "Ontological Relativity."[54] Here, Quine argues that questions of identity and difference make sense only relative to some background language. The account of identity presented here is connected with Quine's thesis of ontological relativity. Essentially, that thesis is the claim that there is no such thing as specifying in any absolute way the items comprising the ontology of a language. The best we can do, Quine claims, is to identify those items relative to the referential base of a background language. But, according to Quine, "Identity is of a piece with ontology." If we can specify the ontology of a theory only relative to some background language, then it is impossible to speak of the identity of objects, except as relative to some chosen coordinate system. Quine illustrates the relativity of identity with an example that is, in some respects, like Geach's. If a branch of economic theory, which does not distinguish between people of the same income, is taken as our background language, then, Quine claims, we must construe persons who are, in terms of a background language with a richer ideology, numerically different, as identical. Like Feldman, Quine finds in this sort of example no reason for rejecting Leibnizian identity. Indeed, Quine defends both Leibniz' Law and the Identity of Indiscernibles (in the sort of amended form Geach presents). Although judgments of identity are always relative to some background language, they are always made in conformity with the concept of Leibnizian identity; for according to Quine, to say that an object, *x*, and an object, *y*, are identical, is just to say that *x* and *y* are indiscernible with respect to the descriptive resources of that language.

D. Strawson on Particulars

Unquestionably the most influential treatment of particulars to appear in recent years is found in the first half of Strawson's *Individuals*. For Strawson, the notion of a particular is extremely broad; it includes not only material bodies and persons, but also things as diverse as states of consciousness and physical processes. His analysis of particulars focuses on what he calls particular-identification. He is concerned with the machinery our conceptual scheme affords us for making uniquely identifying references to particulars.[55]

In the most basic kind of case, Strawson claims, we refer to a particular in our immediate perceptual environment. Since the particular is within the

[54] First appeared in *The Journal of Philosophy*, vol. 65 (1968), pp. 185–212. It is the second chapter of *Ontological Relativity*, *op. cit.*

[55] For discussions of topics central to Strawson's account, see D. Pears, "A Critical Study of P. F. Strawson's *Individuals*," *The Philosophical Quarterly*, vol. 11 (1961), pp. 172–185 and pp. 262–277; Julius Moravscik, "Strawson on Ontological Priority," *Analytical Philosophy*, ed. by R. Butler (Oxford, 1965); Michael Woods, "Identity and Individuation," *Analytical Philosophy* (*ibid.*); David Wiggins, "The Individuation of Things and Places," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVII (1963), pp. 171–202; Michael Woods, "The Individuation of Things and Places," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVII (1963), pp. 203–216; B. A. O. Williams, "Mr. Strawson on Individuals," *Philosophy*, vol. 36 (1961), pp. 309–332.

range of sensible discrimination, it can be identified demonstratively, by the use, that is, of pointing and the demonstratives "this" and "that." More problematic is the identification of particulars outside this range. According to the Identity of Indiscernibles (or one version of it), the non-demonstrative identification of particulars proceeds by way of purely general descriptions.[56] Strawson rejects this sort of account. No description of a particular in terms of purely general predicates (predicates involving no essential reference to particulars) carries with it any logical guarantee that it applies to just one individual. No matter how detailed the list of purely general predicates, it is always possible, Strawson argues, that the description formed out of those predicates will apply to several different particulars.

Like Kant, Strawson wants to claim that space and time together constitute a system of relations adequate for the purposes of particular-identification. We operate, he tells us, with a unified conceptual scheme in which every particular is related to every other by means of space and time; but a consequence of this is that non-demonstrative identification can always be tied to demonstrative identification. We can always uniquely identify a particular outside our immediate perceptual environment by specifying its spatio-temporal relations to particulars within that environment.

Particulars can also be reidentified; we can, that is, identify a particular as numerically identical with a particular encountered earlier. In discussing reidentification, Strawson is concerned to undermine Humean scepticism about numerical identity through time. We speak of reidentifying particulars even where there is a break in our observation of them; but the fact that our observation of particulars is discontinuous, Strawson argues, provides no support for Humean scepticism. Indeed, it is a condition of our possessing the unified conceptual scheme we do possess that we ascribe numerical identity to objects discontinuously observed. If we never did this, there would be no question of a single unified scheme; we would have instead the notion of many fragmentary, discontinuous schemes, one for each stretch of uninterrupted observation. But if our experience were of this sort, the sceptic's doubts about numerical identity could never arise. For those doubts to be possible, we must have the very sort of unified framework the sceptic's doubts about reidentification call into question. Those doubts, Strawson concludes, cannot be other than incoherent.

Strawson agrees with Aristotle in holding that, among particulars, substances (material bodies and persons) have ontological primacy. In Strawson's account, this primacy exhibits itself in the role substances play in particular-identification. Substances, we are told, are the basic particulars of our conceptual scheme: the identification of substances need not depend upon the identification of non-substantial particulars; but were we unable to identify substances we could not make all of the identifying references we do make to particulars of other categories. In support of the claim that material bodies and persons are basic particulars, Strawson argues as follows:

> It seems that we can construct an argument from the premise that identification rests ultimately on location in a unitary spatio-temporal framework of four dimensions, to the conclusion that a certain class of particulars is basic in the sense I have explained. For that framework is not something extraneous to the objects in reality of which we speak. If we ask what constitutes the framework, we must look to those objects themselves, or some among them. But not every category of particular objects we recognize is competent to constitute such a framework. The only objects which can constitute it are those which can confer upon it its own, fundamental characteristics. That is to say, they must be three-dimensional objects that endure through time. They must also be accessible to such means of observation as we have; and since those means are strictly limited in power, they must collectively have enough diversity, richness, stability, and endurance to make possible and natural just that conception of a single unitary framework which we possess. Of the categories of objects which we recognize, only those satisfy these criteria which are, or possess, material bodies—in a broad sense of the expression. (Pp. 28–29.)

Although Strawson thinks that this argument is sound, he finds it "too vague and general." To supplement the argument, he works out a detailed account of the ways in which the identification of substances is involved in the identification of particulars of other sorts. There is, he contends, no general schema exhibiting the relevant identifiability-dependence. In some cases, the dependence is very direct. Thus, we can identify an individual sensation or mental event only by referring to the person whose sensation or mental event it is. In other cases, the dependence is much weaker. We can, for example, refer to a strike or a lock-out without referring to any material bodies or persons; nevertheless, Strawson argues, our ability, in general, to identify things like strikes and lock-outs presupposes our ability to make identifying references to things like persons, tools, and factories. In the case of still other non-substantial particulars, there is no dependence at all. Thus, Strawson is willing to grant the possibility of constructing a series of flashes and bangs such that the identification of the members of the series would, in no way, presuppose the ability to identify substances. What he is concerned to deny, however, is that particulars of this sort could function as the basic particulars of a conceptual scheme, in relevant respects, like ours.[57]

---

[56] The Identity of Indiscernibles is a many-faceted principle; or better, there are a number of related principles that go by that name. For a discussion of the various formulations and uses of the principle, see Loux's "Particulars and Their Individuation," *op. cit.* For further discussion of the Identity of Indiscernibles and related issues, see V. C. Chappell, "Sameness and Change," *The Philosophical Review*, vol. 69 (1960), pp. 351–362; Fred Feldman, "Leibniz and Leibniz's Law," *The Philosophical Review*, vol. 79 (1970), pp. 510–522; Robert Muehlmann, "Russell and Wittgenstein on Identity," *The Philosophical Quarterly*, vol. 19 (1969), pp. 221–230; Douglas Odegard, "Indiscernibles," *The Philosophical Quarterly*, vol. 14 (1964), pp. 204–213.

Strawson claims that his account is a case of descriptive rather than revisionary metaphysics; it is an attempt "to describe the actual structure of our thought about the world" rather than an attempt "to produce a better structure."[58] While I tend to be sympathetic with many of Strawson's substantive claims, I find this distinction puzzling. Strawson suggests that a large number of historical metaphysicians have been revisionists. Admittedly, some philosophers (Parmenides, for example) would want to label their work revisionary in Strawson's sense; but these philosophers have doubtless been in the minority. Philosophers like Descartes, Leibniz, and Berkeley (all of whom Strawson classifies as revisionists) would probably object to the label. How could one show them wrong here? Is it that their systems are bizarre? But what is to count as bizarre in a philosophical system? Possibly Strawson would say that their accounts are incompatible with the layman's picture of the world; but this is to imply that the layman has a properly philosophical picture, a picture that is at the same level as and can, therefore, compete with the philosopher's account. Surely, this is a dubious claim.

The difficulty here is that, by almost any reasonable criterion, one man's descriptivist will turn out to be another man's revisionist. The use of these labels presupposes and is relative to what we construe to be the correct account of the "actual structure of our thought about the world." But, then, of course, Strawson's distinction has, at most, polemical value; and Strawson does, in fact, use it as a polemical tool. There are contexts (luckily, few) where Strawson gives the impression that to refute a view, it is sufficient to show it revisionary in this or that detail.

Suppose, however, that we can make some minimal sense of the distinction. Why is it that a system of revisionary metaphysics is *prima facie* objectionable? In *Individuals* Strawson claims to be providing an account of the most general features of our thought about the world, an account of that "massive central core of our thinking which knows no history." But is it not at least conceivable that a revisionary account would be appropriate here? Is it not at least in principle possible that there be sound philosophical reasons for supposing that the "massive central core of our thinking" is inadequate, confused, or even inconsistent? Strawson evidently thinks not. As it turns out, these methodological claims do not play a central role in Strawson's arguments; and that, I think, is fortunate.

University of Notre Dame

Received March 31, 1971

---

[57] For an attempt to develop, somewhat along Strawsonian lines, the primacy of material bodies, see W. D. Joske, *Material Objects* (London, 1967).

[58] For further discussions of this and other topics central to *Individuals*, see Strawson's book on Kant, *The Bounds of Sense* (London, 1967).

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 2, April 1972

II. INTENTIONS AND INTENDING

HECTOR-NERI CASTANEDA

THIS is an essay on *practical* intentions and *practical* intendings. I shall, however, for the most part leave the italicized adjective implicit. The essay deals with intentions and intendings in the sense in which a man is said to have good (or bad) intentions, or to perform acts intentionally or with a certain intention, or it is said that he merely intends to do an action which he is not performing at the time in question.[1] This is not a proto-philosophical essay intent in gathering philosophical data by means of careful examination of the different senses or uses of the variegated locutions used to attribute intentions or attitudes of intending to different persons. It is a fully philosophical essay that tries to formulate a theory that answers three questions about the nature of intention and intending. Let me explain my three questions:

Consider statements made by uttering the sentence

(1) John intends to come.[1a]

Such statements attribute to John a psychological state or attitude which I shall call *intending*. But those statements present a certain thought content, a *noema* as I shall say, which is the content, or "object," if you wish, of John's intending. This noema is formulated *in* sentence (1) either by the infinitive phrase "John . . . to come" or just by the infinitive "to come." In any case, the full content (whatever it may be at this juncture) will be called an *intention*. This terminological decision parallels the decision often made to speak of a state (even an act) of *believing* and a content which is called *belief*. Needless to say, the word "intention" is often used to refer to the state of intending, just as the word "belief" is often used to refer to the state of believing. When we say that a man's belief is true, we are, of course, referring to the content, i.e., a proposition or a statement. And we do very much the same when we say that a man's intention is good: we are there referring to the content of the man's intending. But perhaps we do refer to his intending itself.[2] In any case, the ambiguity must be dissolved, and my terminology dissolves it.

Now I can formulate my three problems: (A) What exactly are intentions? More specifically: Are they, like beliefs, propositions, or are they acts (actions), or something else? (B) What is the fundamental implicational structure of intentions? Is it the same as, or parallel to, the implicational structure of propositions? (C) What is the fundamental implicational network in which intendings are implanted? Is it the same as, or parallel to, that in which believings are implanted?

Obviously, these logico-ontological questions are crucial to the philosophy of action or practice and lie at the very foundation of moral philosophy. Naturally, any theory that answers all three questions must be a nuclear part of a general theory of practical thinking, or the language of conduct, if you wish. This is the case with the theory of intentions I shall develop in this essay. It is part of a general theory of practical thinking that I have been expounding in a slow moving train of studies.[3] It was first adumbrated in print in 1960.[4]

---

[1] Practical intentions are, obviously, contents (or "objects," if you prefer) of consciousness, more specifically, of the consciousness realized in practical thinking. Thus, they are special cases of intentions in the more general sense in which by Brentano's thesis of intentionality every mental act or attitude intends an intention.

[1a] As George Pitcher reminded me, things other than one's actions are said to be intended. We say, for instance, "Jones intends that it rain" and "Jones intends Mary to come." They are not studied here. In my views they are not intentions in the same sense as one's intentions to do something. See Section VII below.

[2] For an analysis of the goodness of intentions see my "Goodness, Intentions, and Propositions," in Paul Welsh (ed.), *Festschrift to Charles Baylis* (forthcoming).

[3] Besides the essays mentioned in footnote 2 above I should mention: "Imperatives, Decisions, and Oughts: A Logico-metaphysical Investigation" in H.-N. Castañeda and George Nakhnikian (eds.), *Morality and the Language of Conduct* (Detroit, Michigan, 1963), pp. 219–300; paperback, 1967), pp. 219–299; "Actions, Imperatives, and Obligations," *Proceedings of the Aristotelian Society*, vol. 68 (1967), pp. 25–48; and "On the Semantics of the Ought-to-do," *Synthese*, vol. 21 (1970), pp. 449–468.

[4] In my "Outline of a Theory on the General Logical Structure of the Language of Action," *Theoria*, vol. 26 (1960), pp. 151–182, and in the appendix to my "Imperative Reasonings," *Philosophy and Phenomenological Research*, vol. 21 (1960), pp. 21–49

My answers are, roughly, these: (A.1) intentions are neither propositions nor actions; (A.2) intentions are proposition-like, and in the case of sentence (1) above, the intention is expressed by the infinitive "John . . . to come"; (A.3) intentions are first-person prescriptions, for the explanation of which see Section IV below; (B.1) the structure of intentions is entirely parallel to the structure of propositions; (C.1) the structure of intending is like that of believing, but not quite parallel in that principles of the heredity of intending depend on believing. In addition to the theory itself, my main contributions here are the arguments for theses (A.1) and (B.1).

To furnish some motivation for the theory to be developed, I shall consider some important data that have emerged from a recent symposium on the structure of intention, i.e., intending.

I. Some Crucial Data

Intentions have recently received a lot of attention. A large amount of excellent data have been gathered and used in efforts at mounting theories on them as well as in attempts at toppling such theories. Both of these activities are masterfully illustrated in the three superb papers on the structure of intending by Roderick Chisholm, Annette Baier, and George Pitcher, respectively, presented in a 1970 symposium of the Eastern Division of the American Philosophical Association.[5] The data stored in these papers, especially in Annette Baier's, simply have to be accounted for by any adequate theory of intention.

The symposium starts with Chisholm's attempt at analyzing a cluster of concepts which center around the idea of intending. Chisholm's primitive concept is neither that of intending nor that of intention. It is the concept *x brings it about that p in intending to bring it about that q*. Both Baier and Pitcher take him to task for adopting as primitive a concept which from every point of view seems to be rather complex: not only does it seem to have as constituents the concepts of *bringing about* and *intending*, but also these are connected by a perplexing structure expressed with the word "in." Baier and Pitcher take issue with Chisholm on several other claims as well.

Against the foil of Chisholm's analysis his critics' discussions do secure some important phenomenological data which are criteria of adequacy for any theory of intentions. The most important of such criteria are:

D1. There must be a fundamental theory of intention that deals with the concept *x intends A*, bereft of superstructures and connections, e.g., like those surrounding it in the case of Chisholm's "primitive" concept *x brings it about that p in intending to bring it about that q*. (A. Baier, 650ff; Pitcher, 659ff.)

D2. Intentions are essentially and unavoidably first-personal. (A. Baier, 658.)

D3. The concept *intending to bring about* presupposes the concept *intending to do*, and must be analyzed in terms of it. (A. Baier, 656ff.)

D4. The "objects" of intending are not states of affairs. (A. Baier, 652, 657f.)

D5. Intentions are not goals or purposes. (A. Baier, 651ff.)

The symposium contains interesting discussions on two principles in the logic of intending.[6]

These are proto-philosophical data which it pleases me to find satisfied, clarified, and organized in my fundamental theory of intention and intending. I call it "fundamental" because it deals with the most basic laws governing the structure of intentions and intending. Analyses of the more complex concepts that concern Chisholm must be superimposed on the fundamental theory. At that level other criticisms of Chisholm's proposed analyses made by Baier and Pitcher must be examined. However, this will not be attempted in this essay.

II. Intentions, Actions, and Propositions

Since we are not dealing with the contents of theoretical reason, we shall simply assume that the terms "state of affairs," "proposition," and "statement" refer to the same entity. The first term is the most neutral and objective one; the second term puts a slant on the entities in question being possible noemata or "contents" of psychological acts or attitudes; the term "statement" puts the slant on those entities being the contents of linguistic acts. Similarly, the words "beliefs," "suppositions," "conjectures," and "assumptions" refer to propositions *qua* noemata of specified acts or attitudes, namely, believing, supposing, conjecturing, and assuming, respectively. We shall leave it open here whether or not a true proposition, viz., an existing state of affairs, is identical with a fact. And we shall not preclude that all of these entities can in last, or penultimate, analysis be reducible to sentences or to behavior involving sentences. By adopting a Platonistic-looking framework we can make distinctions without begging questions in the metaphysics of abstract entities: we shall be at least formulating phenomenological distinctions that should be taken as criteria for the adequacy of any reductionistic metaphysics to be considered later on: the theory of the structure of believing (or intending) and its phenomenological contents is at least a *macro* theory to be explained or accounted for by a nominalistic *micro* theory.

---

[5] These papers are all printed together in *The Journal of Philosophy*, vol. 67 (1970), as follows: Roderick Chisholm's "The Structure of Intention," pp. 633–647; Annette C. Baier's "Act and Intent," pp. 648–658; and George Pitcher's " 'In Intending' and Side Effects," pp. 659–668. I examined this symposium in "Intentions and the Structure of Intendings," *The Journal of Philosophy* (forthcoming). Valuable data are also found in Jack W. Meiland, *The Nature of Intention* (London, 1970).

[6] See my discussion of such principles in the paper mentioned in footnote 5.

Now, whether intentions are propositions or not, sentences of the form "*X* intends to do *A*" are, normally, expressive of propositions: namely, psychological propositions that appear to consist of a psychological relationship, intending, holding between an agent and an action: the agent being represented by '*A*'. Yet this is only a grammatical appearance. For, clearly, the action '*A*' stands for in "*X* intends to *A*" is a universal, an action which typically can be performed by many agents. Yet the thought content before *X*'s consciousness cannot be the universal action: his intending to do *A* is his determination to *do A himself*, i.e., the particularized action consisting of the action *A as done by him*. An intention is thus a complex like a proposition in that it has a subject, namely, the agent conceived in the first person, and a predicate, namely, the general action *A*. Thus, the propositions formulated by sentences of the form "*X* intends to do *A*" are philosophically more perspicuously expressed by sentences of the form "*X* intends that he do *A*." The question that surges at this juncture is whether or not an intention is a proposition. I shall argue that it is not. But first I want to reply to Annette Baier's arguments in defense of the view that the full content of an intending is an action.[7]

Baier insists that intentions are first-personal, yet she contends that intentions are actions:

> Now it appears that . . . [there is] a dummy agent, or rather agent-patient, since the "I" who is intending always figures in his own intentions. Is this objectionable, or is it a mark of that necessary truth mentioned earlier, that I can intend only *my* doings? . . . if I intend to speak, this is not to intend me to speak . . . we drop off our intentions [i.e., intendings], they are simply intentions [i.e., intendings] to do. (*Op. cit.*, p. 658.)

Baier's conclusion seems to be that the noema, or what is before a person's mind, when he comes to intend to speak is simply the action to speak. Yet she claims that one can only intend *one's* doings, and this implies that the idea of oneself as agent is part of the intention. I find this difficult to reconcile. The reason she gives is that the sentence "I intend to speak" is not in ordinary language an abbreviation of "I intend me to speak." This linguistic point is, it seems to me, absolutely correct. Yet this grammatical fact cannot establish that one does not have to think of oneself as agent to intend to do something. Consider, for proper perspective, that the sentence "I shaved" is not in ordinary English an abbreviation of the sentence "I shaved me"; yet this fact does not have the tendency to show that shaving is not a dyadic relation which is reflexively instantiated when a person shaves.

Now, a man who intends to speak may very well express his intention, i.e., the full content of his intending, by saying "I shall [or will, depending on his dialect] speak." Thus, we have here in oratio recta that the first-person reference is part and parcel of a man's intention. Therefore, even the linguistic data are ambiguous: in "John intends to speak" there is no explicit first-person reference as part of what John intends, but there *is* in his asseveration "I will speak." Actually, this conflict in the data is not an impasse. It is a "conflict" between a piece of oratio obliqua in a larger context, where elisions of iterations are natural, and a whole oratio recta. Undoubtedly, we must adopt the principle that the oratio recta is more revealing than its corresponding oratio obliqua, of the structure of what they both express. Hence, I regard this as conclusively nailing down the view that an intention is a proposition-like structure which in the simplest case has as subject an agent conceived in the first-person way and has an action as predicate.

At this juncture it is tempting to suppose that intentions are propositions and that the difference between a man intending to speak and the same man believing that in fact he will speak is not a difference in noemata, but a difference in psychological states or operations.[8] It might be supposed that in this case the believed noema and the intended noema are identical, namely, the first-person propositions the man in question would express by saying "I will speak." I proceed to argue that this view is downright mistaken.

---

[7] Meiland also holds that the objects of intention (i.e., intendings) are actions (*op. cit.*, p. 43).

Intentions are not their corresponding propositions. This is apparent from reflection on the structure of complex intentions that have as components both propositions and intentions. Consider, for instance,

(2) After careful deliberation Jones arrived at the following intention: both to drive his daughter to the shopping center if she comes early and to buy his wife a purse if he drives his daughter to the shopping center.

Since Jones's intention is somewhat complex we need the grouping or scoping device "at the following intention" to indicate clearly that *everything* said after the colon is part of Jones's intention. From the preceding discussion we know that the full intention includes Jones's first-person references as the subjects of the actions of driving and buying. Thus, in pedantic, but philosophically more perspicuous English, we can put statement (2) as follows:

(2a) After careful deliberation Jones arrived at the following intention: both that he-drive-his-daughter-to-the-shopping-center if she comes early and that he buy his wife a purse if *he drives his daughter to the shopping center.*

Let us examine the structure of Jones's intention. Obviously, it is a conjunctive one, and each conjunct is a conditional intention. Each of the conditional component intentions is, however, a mixed intention: the antecedents are propositions, while the consequents are the simplest intentions that enter into the total complex intention of Jones's. I have chosen the example so as to highlight the difference between the intention expressed by the hyphenated subjunctive clause and the proposition expressed by the indicative clause in italics. Here we have an intention and a proposition that structurally correspond to each other in that they have the same subject and the same predicate. But they are different.

If the intention expressed in (2a) by "he-drives-his-daughter-to-the-shopping center" were the proposition (or state of affairs) expressed by "he drives his daughter to the shopping center," then it would follow from both (2a) and Jones applies hypothetical syllogism to (2a) if and only if it is applicable that

(3) Jones intends the following: that he buy his wife a purse if his daughter comes early.

Since it may be thought that (2a) implies (3) in any case, it is all the more worthwhile to show that this is not so. One way of seeing it is perhaps this. Suppose that Jones is very prone not to carry out his intentions. Thus, he has arranged with each member of his family that the latter will punish him in some way when he breaks an intention of his that involves that member of his family. This is the reason why he is so deliberate in framing his intention formulated in (2a): by it his wife is the beneficiary of a conditional intention whose condition is that he drives his daughter, not that his daughter comes early. Thus, when it happens that his daughter comes early and he refuses to drive her to the shopping center, he is liable to be punished by his daughter; but since his intention benefiting his wife has an unfulfilled condition, his not buying a purse for his wife does not make him liable to be punished by his wife. On the other hand, if (3) were a consequence of (2a), failing to carry out his intention benefiting his daughter would not relieve him of his liability due to his failure to carry out his intention benefiting his wife. Yet this relief is precisely what he wanted to insure by carefully framing intentions so related that the intentional consequent of one is a propositional condition of the other. Since it *must* be possible for Jones to secure that relief, without having to disown any of his intentions, (2a) does not imply (3) by itself or together with condition about Jones's application of hypothetical syllogism.[8a]

Since intendings are mental states on intentions, it must be the case that the network of implications governing intendings is supported by the network of implications governing intentions. In particular, it is natural to suppose that a principle like the following is part of the support the latter network provides to the former:

(*I) "[X is rational with respect to his intentions, and] X intends [intention] i and X

---

[8] Chisholm appears to take the view that the items which are involved in, or are contents of, intentions are state of affairs. (*Op. cit.*, pp. 635ff.)

[8a] I am indebted to George Pitcher for this condition.

> believes that *i* implies *j* [a possible intention of *X*'s]" implies "*X* intends *j*."

I have deliberately formulated (*I) so as to be neutral with respect to the ontological status of intentions. Thus, adopting (*I) we can argue that intentions cannot be mere actions or propositions. The conjunct within brackets is thought by some philosophers to be deletable. I will delete it in the rest of this section, but the reader may supply it if he wishes.

If intentions were simply actions, or if they were identical with their corresponding propositions, (*I) would be the same as

> (I.a) "*X* intends to *A* and *X* believes that his *A*ing implies his *B*ing" implies "*X* intends to *B*."

In fact, this looks like a nice principle, even if one holds that intentions are not actions. Thus, showing that (I.a) is false is a very fruitful thing all around. Consider, therefore, the case of our friend Corliss who knows that he is driven to hurt the feelings of people near him when there is a full moon. His friend Marco will visit him from April 1 through April 15, 1971. Corliss is worried about his behavior on April 10. Very deliberately Corliss frames the intention to do something special for Marco on April 11 when he recognizes that he has unintentionally hurt Marco's feelings on April 10. Clearly, Corliss' doing something special for Marco on April 11 when he recognizes that he has unintentionally hurt Marco's feelings on April 10 *implies* Corliss' (unintentionally) hurting Marco's feelings on April 10. Yet, though he intends the implicans he does not intend the implicate. Therefore, (I.a) is false. Furthermore, if (*I) holds, Corliss' intention is neither the action doing something special for Marco when he recognizes that he has hurt Marco's feelings, nor the proposition that Corliss does something special for Marco on the 11th when he recognizes that he has unintentionally hurt Marco's feelings on the 10th.

It may not be amiss to point out that the Corliss example is also a counter-example to revisions of (I.a) that include any combination of the following conditions on action *B*: (i) the agent *X* can perform *B* (ii) *X* can initiate his doing *B* (iii) *X* can intentionally perform *B* (iv) *X* does not intend not to do *B*. All conditions (i)–(iv) are satisfied by Corliss and his hurting Marco's feelings. So, if (*I) is true, these conditions, added to the proposition Corliss would express by saying "I will hurt Marco's feelings," do not suffice to produce a propositional analysis of the intention pertaining to, but not endorsed or adopted by, Corliss: to hurt Marco's feelings.

In the Corliss example the action of hurting Marco's feelings enters in Corliss' intention considered as a *circumstance*, not as itself material of intention. Similarly, I shall say that the proposition "Corliss will hurt Marco's feelings" enters in Corliss' intention as a *circumstance*. My contention is, thus, that an action can be part of an intention only if it enters in the latter in some special way, a way characteristic of the basic noemata of practical thinking, and, hence, not to be found in propositions. But of this more later.

It may be thought that an intention is a proposition and that the state of intending is more perspicuously expressed by the locutions "intends to bring it about that" or "intention to bring it about that." Thus, it may be contended that

(4) Jones's intention is to open the door

should be understood as:

(5) Jones intends [or his intention is] to bring it about that he opens the door.

No doubt, we can use the two sentences to make exactly the same statement. Yet, given the primary semantical rules governing them, they can be used properly to formulate different propositions. The difference I have in mind is essentially the difference illustrated above in the Corliss example, but now the situation is more delicate.

In the Corliss example Corliss' action of hurting Marco's feelings is a pre-future, or future-perfect circumstance within Corliss' intention to do something special for Marco. In the case of Jones in (5), Jones's action of opening the door in question is a post-future circumstance within Jones's intention. According to (5) Jones intends to do some action, which is not specified, but which fits in with his other purposes and goals, so that as a consequence of that action of his he will be in a state that will cause his opening of the door. Patently, Jones may cause his opening of the door by persuading a hypnotist to put him in a hypnotic trance and order him to open the door, or to perform some other action which will inevitably lead Corliss to open the door unintentionally and unawares. Clearly, Jones's causing that it be the case that he opens the door is different from Jones's opening the door: he can do the former without *in* doing it opening the door. Thus, his intending the former, which is what he intends when he intends to bring it about that he opens the door, is not the same as his intending to open the door.

I conclude, therefore, that *bringing-about* is not a feature characteristic of the state of intending, but an element that enters as a constituent of *some* intentions. Furthermore, an intention "I shall do *A*" is different from the corresponding intention "I shall bring it about that I do *A*." Finally, an intention "I shall do *A*" is not identical with its corresponding proposition "I will bring it about that I (will) do *A*."

At this juncture it may be suggested that intentions are propositions, but not the ones corresponding to them structurally. I, however, have no idea as to how to start looking for a proposition to be identified with a given intention, say, my intention to finish this paper. And there is one general argument against any such identification, namely: Propositions are either true or false, while intentions are neither true nor false. But a full evaluation of this argument requires a detailed account of truth. Since I cannot do this here, I propose to regard this argument not so much as conclusively showing that a theory that treats them as distinct entities is at least worth developing. Nevertheless, philosophers with a reductionistic bent should attempt to reduce intentions to propositions. Philosophical illumination always results from striking two or more theories on the same topic against each other.

III. Intentions

We need, according to the preceding argumentation, a theory of intentions that does not identify them with propositions. And we can in general express this non-overlapping of the two categories of noemata by saying that intensions have no truth-values, while propositions are truth-valued. We have seen that an intention is expressed by the clause "John . . . to come" *within* either of the larger contexts below:

(1) John intends to come;
(1.b) John's intention is to come.

And we have also noted that intentions are often formulated by future-tense sentences of the form "I shall [will] do *A*." Now, it appears entirely natural to think that what John fully intends in the above example is the same as what he would express, had he *decided* to come, by saying "I shall come." I propose to enthronize this natural identification as a *theoretical thesis: the very same entity or content which is a decision, or resolution, is an intention.* Whether an intention is a decision or not, then, depends solely on whether a psychological process of deciding or deliberating has the intention as its ending element. Partial support for this thesis is furnished by the widespread view that decisions are neither true nor false, i.e., are not propositions. Thus, basic intentions have, *inter alia*, the following important features: (i) an intention is a propositional-like content in that it has as constituents both a subject and an attribute which is an action; (ii) like a proposition, a decision or intention does not have the person, who is the agent, himself as a constituent, as can be seen from the fact that in the above example John may fail to believe that he himself is John; (iii) an intention does have as subject constituent the sense (to speak à la Frege) or the individual concept (to speak à la Carnap and Church) of the agent's first-person references; (iv) the attribute constituent of an atomic intention is an action, just as it is in the corresponding performance proposition, but (v) this attribute is linked to its subject, the agent concept, by means of a *different copula.* This intentional copula is characteristic of the autonomy of practical reason. (The intentional copula can be further analyzed as an operator on actions—in accordance with natural [set-theoretical] semantical models for a calculus of intentions and propositions.)

One comment on (iii) is due. An intention, or decision, is never a third-person content. (Of course, a man can use a third-person sentence when he means to be expressing first-person contents.) This is not the place to argue for the primitivity, i.e., the irreducibility of propositions that contain first-person references, i.e., those which constitute self-knowledge and self-belief, to second- or third-person propositions. I have argued for this at length elsewhere.[9] Fortunately, we can leave this in abeyance here. Whether reducible or not, there is a *prima facie* distinction between first- and third- or second-person contents (propositions, intentions, etc.).

---

[9] More directly in "On the Phenomeno-logic of the I" in *Proceedings of the XIVth International Congress of Philosophy*, vol. 3, (Vienna, Austria: Herder, 1969), pp. 260–266, which is an introduction to my tract on demonstrative reference both in *oratio recta* and in *oratio obliqua* "Indicators and Quasi-Indicators," *American Philosophical Quarterly*, vol. 4 (1967), pp. 85–100; see also my "On the Logic of Attributions of Self-knowledge to Others," *The Journal of Philosophy*, vol. 65 (1968), pp. 439–456.

At any rate, we have two crucial contrasts on our hands:

(A) The contrast between the two propositions:
  (1) John intends to come;
  (2) John believes that he (himself) will come.[10]

(B) The contrast between intention (3) and proposition (4) below, which are the contents of the intending of (1) and the believing of (2):
  (3) I [= John] come
  (4) I [= John] will come.

Note that the parenthetical "[= John]" is an index on the first-person pronoun signaling a comment *by us* on, and external to, intention (3) and proposition (4). This comment is intended to help identify the intention and the proposition in question.

We have briefly dwelt upon contrast (B). Let us say something about contrast (A). Intending is, in the exercise of practico-theoretical reason, the counterpart of believing. Now in the most customary sense of the word, believing is a disposition or propensity which is exercised in acts of thinking-that (such and such is the case). Likewise, intending is a disposition or propensity which is exercised in acts of deciding, and in affirmations or rehearsals of decisions. But unlike believing, intending is not a purely intellectual propensity; it is a propensity that exercises practical reason: it rearranges the motivational or causal routes constitutive at the time in question of the person whose intending it is. But whatever the analysis of the motivational role may be, we must note that the words "think" and "know" are general terms that apply both to purely cognitive and to practico-cognitive states or operations of reason. Thus, both believings and intendings are exercised in acts of thinking. While the former exercises are thinkings-that, the latter exercises are thinkings-to, e.g., "John was thinking to come." Parallelly, one knows that something is the case, and one (sometimes) knows what to do.

In brief, the logico-ontological parsing of

(1) John intends to come

may be represented as:

(1A) John intends to (I [= John] shall) come.

IV. Intentions and Prescriptions

Intentions or decisions (noemata, that is) belong with pieces of advice, requests, orders, commands, and, in general, with the contents of assertive-like and deliberate acts of telling a person what to do. I have used before the word "*mandate*" as the generic term to group all the contents of such acts, and the word "prescription" to refer to the structure common to a family of mandates that demand of the same agents the doing of the same actions in exactly the same circumstances. For example, the order "Peter, go home," the request "Peter, please go home," and the advice "Peter do go home" (assuming, of course, that we have the same Peter and the same time throughout), have as underlying structure the prescription "Peter go home"—expressed clearly by the subjunctive clause. Prescriptions are characterized by: (i) being neither true nor false, and (ii) not being the contents of believings or knowings-that. The atomic ones are constituted by an agent concept tied to an action attribute by means of a special copula. Prescriptions are not necessarily second-person contents. Mandates are, of course, second-person contents. But a prescription like "Peter go home" is a third-person prescription, and a prescription like "You go home" is a second-person prescription. But to issue a prescription is to issue it embellished by, or clothed under, the guise of one or another type of mandate. Issuing a prescription (or mandate) is the linguistic act counterpart of stating a proposition.

In my view prescriptions form a genus under which we find intentions. Specifically, at this juncture my crucial thesis is that *intentions are first-person prescriptions: the intentional copula [or action operator] is exactly the same as the prescriptional copula.* This rules out the indentification of intentions with a subset of mandates. Therefore, intentions are *not* first-person mandates. There is, however, a variant view[11] that finds for the case of intentions a duality like the one there is between mandates and second-person prescriptions. According to this view: (a) decisions and intentions are the first-person counterpart of mandates, and (b) just as mandates are embellishments of prescriptions, intentions and decisions are embellishments of underlying first-person prescriptions. Here, however, I adhere to the former italicized thesis. I do not (any more) see in the first-person case a duality of psycho-linguistic acts on first-person contents analogous to ordering, entreating, advising, requesting, and commanding. These psycho-linguistic acts on prescriptions depend on a social organization which simply cannot have a counterpart in the case of a man who is making decisions. For example, deciding and intending are not alternative ways of handling the same content. Furthermore, an intention a man endorses may be a decision; it seems a purely contingent matter that not all intentions are decisions. But all decisions *are* intentions. Perhaps some orders order the same prescription (or action) a given request requests; but no order *is* at the same time a request made by a person to one and the same agent. In general there is even a sharp difference between the actions a man can order his subordinates to perform and the ones they can request him to do.

---

[10] Note that the pronoun "he (himself)" which appears in *oratio obliqua*, e.g., in a belief context in sentence (2), is used to attribute self-reference to the person referred to by its antecedent "John." It is, in my terminology, a *quasi-indicator*. I have studied it in the papers mentioned in footnote 7 as well as in " 'He': A Study in the Logic of Self-consciousness," *Ratio*, vol. 8 (1966), pp. 130–157, where I contrast it with six other uses of the third-person pronoun. In the case of sentence (1) we must count a quasi-indicator represented by a null word.

[11] I held this view in 1956; there is a trace of it in the *Theoria* paper cited in footnote 4 above.

In any case, the fundamental contrast we are dealing with is not the contrast between intentions and propositions. It is the contrast between prescriptions and propositions. Thus, the operations of the mind divide into those operations addressed to propositions, the purely cognitive or theoretical ones, and those addressed to prescriptions, the practico-cognitive or practico-theoretical ones. The former are operations that map (in the logico-mathematical sense) propositions into propositions; the latter map prescriptions into propositions. This is the general contrast illustrated by examples (1) and (2) above.

V. The Basic Structure of Intentions

Intentions combine with other intentions by means of the regular propositional connectives. There are, thus, denials, conjunctions, conditionalizations, and biconditionalizations of intentions. For example: a negative intention is expressed by "I [= John] won't go home," when this expresses or rehearses a decision; a conjunctive intention is expressed by the subjunctive clause "[John] go home and study" *in* "John intends to go home and study." Some philosophers have argued, and others will undoubtedly argue, that the words "not," "and," "but," and "or" do not express the same connection in examples like the preceding ones. Their reason is that since decisions have no truth-values the English connectives cannot be governed by the ordinary truth-tables when they link sentences or clauses expressing decisions. This argument is in one sense absolutely sound. If the *truth-tables* fully characterize the meaning of "not," "and," and "or," then these words are used in a different sense in the above examples. But so are they when they are used in examples like "Someone is not here" and "Someone is tall or heavy," in which the connectives "not" and "or" express a connection of properties, not propositions. In short, it is an uninteresting verbal matter whether or not intentions are compounded in the same ways propositions are. *A fortiori*, it is also a verbal matter whether intentions (prescriptions) stand in relationships of implication, i.e., the same relationship in which propositions stand. What is important is the fact that intentions parallel propositions both in their way of composition and in their formal structure. I prefer to "extend" the sense of the English connectives and speak of the same forms of composition and of the same rules of implication characterizing each connection or content structure—with just one restriction: to be *compossible* (i.e., disjoinable, conjoinable, etc.) two intentions must have first-person references to the very same agent.[12]

Intentions combine with propositions. For example, John's intention may be "I [= John] will go home before it is dark," where the sentence "it is dark" expresses a proposition, i.e., a truth-valued noema. The combination expressed by the word "before" is a complex involving quantifiers and conjunction: going home at some time t and t being earlier than the time at which it is dark. Thus, intensions and propositions also combine with each other by means of the regular connections. What type of noemata are such mixed combinations? Clearly, the above compound example "feels" like an intention. Likewise, conditional compounds like "If he comes, I will (shall) go home" and disjunctives like "I will go, unless she comes soon" all "feel" intentional. In other cases, however, it is not completely clear how one should classify the complex noema. If a man says "It is getting dark, and I am going home," his token of "and" links a sentence that expresses a proposition and one that expresses an intention; but there is no compulsion to say either that the compound is an intention or that it is a proposition. Doubtlessly, one can develop a theory that allows some compounds to be intentions, others to be propositions, and others to be neither. But the general idea that propositions and, hence, noemata are closed under the logical connections would demand a complicated bookkeeping. Thus, I propose, not as a generalization of what one is inclined to say, but as a theoretical principle, that *mixed connective compounds of intentions and propositions are intentions*. This principle maintains the closure of propositions and of intentions, making propositionhood a recessive character.

---

[12] See my "There are Sh-Command Inferences," *Analysis* (forthcoming), for a turning of the tables on the claim that imperatives are not connected by so-called truth-functional connectives and stand in no implication relationships.

The preceding terminological decisions and principles allow intentions to have a simple basic implicational structure, namely, the one captured by the principles of implication given below, which exhibit the corresponding principles of composition. To save space let A be a set of noemata, i.e., propositions and/or intentions; let 'i' and 'j' represent intentions; let 'X' and 'Y' be any noemata, whether propositions or intentions; let "not-X" represent the negation of noema X, and let "(X & Y)" represent the conjunction of noemata X and Y. For convenience I will say that noemata as well as sets of noemata imply. Thus, the basic implicational structure of intentions is, assuming compossibility, characterized by:

PoI.1. If i is a member of A, then A implies i.

PoI.2. If A implies i, then A implies not-not-i.
[We have as a special case that "I [= John] shall P" implies "I [= John] shall not fail to P."]

PoI.3. If A implies not-i, then A implies not-(i & X).
[Special case: whatever implies "I [= John] shall not P" implies: "I [= John] shall not do the following: P while it is the case that p."]

PoI.4. If A implies not-i, then A implies not-(X & i).
[Special case: whatever implies "I [= John] shall not P" implies "I [= John] shall not both Q and P."]

PoI.5. If A implies i, and A also implies j, then A implies (i & j).

PoI.6. If A and i imply j, and A and not-i also imply j, then A by itself implies j.

PoI.7a. If A implies not-(X & not-i), then A and X imply i.

PoI.7b. If A implies not-(i & not-X), then A and i imply X.

These principles characterize the most elementary structure of negation and conjunction. We leave it to the reader to satisfy himself by the examination of examples that PoI.1–PoI.7b all hold for some class A of noemata that contains intentions only or intentions and propositions. We assume that the other connections relate to negation and conjunction in the customarily assumed way.

By a trivial generalization of the completeness of the standard propositional calculus adopted to the two-sorted calculus constituted by the above eight principles of implication, the latter calculus determines a two-valued structure for intentions. Naturally, there are for intentions value gaps analogous to the truth-value gaps for propositions. But we should be able to adopt similar regimentations. In any case there is the problem of the analysis of such intentional values. Obviously, there must be a designated value which is the counterpart of truth. We do not call it "truth," for it necessarily lacks the requisite correspondence with fact. What are, then, the prescriptional (and intentional) values involved in ordinary practical inference is a hard and very interesting problem. It is part and parcel of the general problem of how intentions connect with desires, goals, purposes, in short, with the motivational nature of an agent. I simply must beg off discussing it here.[13]

Principles PoI.1–PoI.7 are, obviously, principles that govern propositions and intentions. They say absolutely nothing about the linguistic mechanisms employed or employable to express negation or conjunction. This is of utmost importance. The fact that a sentence having a future-tense clause after the word "if" cannot grammatically express a decision, for instance, does not show that there are no pure conditionalizations of intentions. If conditionalization is reducible to negation of a conjunction, then pure conditionalizations of intentions are expressible in ways that do not include the use of "if" or "only if." If conditionalization is a primitive connection, not reducible to negation of a conjunction, then we can at most say that certain compounds are not expressible in ordinary language. But from this it does not follow that they cannot be thought, and much less that they do not occupy their positions in the logical space of intentions and propositions.

---

[13] A preliminary effort, mistaken in detail, appears in "Imperative Reasonings" cited in footnote 4 above. There are also some relevant remarks toward the end of "On the Semantics of the Ought-to-do," cited in footnote 3 above.

VI. The Structure of Intending

As noted above, intending is an operation of the mind on intentions. More specifically, intending is an operator or connection relativized to agents that transforms intentions into propositions. Thus, the basic structure of intending is a two-sorted modal structure. Since intending is the practico-theoretical counterpart of believing, the modal structure of intending is essentially like the structure of believing. This is a simplifying result of our previous theses.

Naturally, there are some complications. For instance, the case of mixed intentions demands a joint treatment of intending and believing. Nevertheless, since intending and believing are special cases of a general operation of the mind that we may call *endorsement of thought contents*, their joint treatment will yield parallel structural properties.

Evidently, many a discussion on the logic of belief, i.e., believing, applies, *mutatis mutandis*, to intending. In particular, this holds for discussions about inconsistent beliefs and quantification into opaque contexts. Here, however, we cannot pursue further the problem of characterizing in general the structure of intending. I shall be content with formulating some principles governing the heredity of intending.

In what follows I shall use formulas without any quotation marks. They must be assumed to have Quine's corner quotes *implicit*.

The fundamental principle of composition governing intending is this: if i is an intention with first-person reference(s) to agent X, then an atomic or basic *intention proposition* is of the form I_x (i), which abbreviates "X intends to do A," where intention i is the intention of the form "I $[= X]$ do A." At this juncture a discussion of what I have called quasi-indicators becomes relevant, for first-person references are attributed, captured but not made, by means of quasi-indicators.[14]

A general principle that seems clear for intending is this:

PoIN.1. I_x $(i \;\&\; j)$ implies I_x (i).

Now, from the above principles PoI.1.–PoI.7 it follows that:

PoI.11. A proposition p and an intention i imply the conjunctive intentions $(p \;\&\; i)$ and $(i \;\&\; p)$.

This principle underlies a brother of PoIN.1, where "B_x $(\ldots)$" is an abbreviation of "x believes that . . .":

PoIN.2. I_x $(i \;\&\; p)$ implies I_x (i) & B_x (p).

The converse of PoIN.1 and PoIN.2 are doubtful as principles of actual intending, i.e., principles not about what a person is committed to intend but about what a person actually intends. We all know how people get into inconsistent beliefs, and there is reason to suspect that they may easily get also into inconsistent intentions. But if a man is rational, at least at a time when he is paying attention to his intentions and his beliefs, it does not seem outrageous to suppose that he may be able to conjoin his beliefs and his intention. Let us abbreviate with "$R(x)$" the proposition that person x is thinking rationally at a given time when he is paying attention both to his intentions involving action A and to his beliefs about the possible effects of his doing A, and let us understand our operators 'I_x' and 'B_x' to denote x's intendings and x's believing at the time in question. Thus, I adopt:

PoIN.3. $R(x)$ & $I_x(i)$ & $I_x(j)$ implies $I_x(i \;\&\; j)$;
PoIN.4. $R(x)$ & $I_x(i)$ & $B_x(p)$ imples $I_x(i \;\&\; p)$;
PoIN.5. $R(x)$ & $I_x(i)$ & $B_x(i$ implies $j)$ implies $I_x(j)$;
PoIN.6. $R(x)$ & $B_x(p)$ & $I_x($if p, then $i)$ implies $I_x(i)$.

Of these principles probably PoIN.3 and PoIN.6 are the least controversial. It must be noted that PoIN.5 is precisely the principle (*I) of Section II above.

VII. Adequacy of the Theory

It is easy to see that the fundamental theory of intending and intention expounded above satisfies the data listed in Section I very well. It is made to order for data D1, D2, D3, and D4. The theory as given includes no analysis of the concepts *bringing about* and *intending to bring about*. But it does provide the prerequisite framework for such analyses. The theory as given also fits datum D5, but it does not contain analyses of *goal*, *end*, *want*, or *purpose*. Suffice it here to indicate that although the locutions of the form "*X*'s goal [end, purpose] is to *A*" and "*X* wants to *A*" do sometimes express intention propositions; more often than not they express other types of propositions. In particular, "*X* wants that *p*" and "*X*'s goal [end, purpose] is that *p* be the case" express propositions resulting from the application of an operator, not to an intention, but to another *proposition*. Consideration of examples should persuade the reader of the validity of the principles listed in Sections V and VI. Principle PoIN.6 is defended in Section II, and elsewhere[15] I defend both it and PoIN.4.

---

[14] See the papers cited in footnotes 9 and 10 above.

Propositions expressed with sentences of the form "*X* intends *Y* to *A*" are of either of two types. On the one hand, they are propositions more perspicuously expressed by the corresponding sentences of the forms "*X* intends to try to make [help] *Y* to *A*" or "*X* intends to make *Y* to *A*." On the other hand, they are propositions that are made up by an operation expressed by "*X* intends" on the *prescription* "*Y* to *A*," not on the intention "I[=*X*] shall *A*" or the proposition "*Y* will *A*."

In short, the theory is certainly as adequate as far as data D1–D7 is concerned. Its extensions should be developed so as to satisfy the more complex data secured by the Chisholm-Baier-Pitcher symposium.

Indiana University

Received March 31, 1971

---

[15] See my discussion of the Chisholm-Baier-Pitcher symposium, mentioned in footnote 5 above.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 2, April 1972

III. AFTER-IMAGES

O. R. JONES

IT is common, in discussions of sense perception, to invoke a broad dichotomy of some kind to help solve problems about the objects of perception. Some philosophers distinguish between what they call "immediate or direct perception" on the one hand, and "perception" on the other. Others speak in terms of sense data and physical objects. More recently, Miss Anscombe has distinguished between intentional and material objects of perception. It seems to me that one can be easily bewitched by the habit of generalizing in terms of such broad distinctions as these with the result that a variety of likenesses and differences between the various objects of perception escape notice. I think that this can be shown to have happened in the case of after-images.

I

I begin with some arguments put forward by Norman Malcolm in his article called "Direct Perception."[1] Malcolm follows Moore in taking after-images to be objects of direct, immediate perception which are thought to contrast with physical objects. Moore's classification of after-images with toothaches under the title "sense data"is echoed in Malcolm's discussion where he compares assertions about an after-image with assertions about a pain in one's shoulder. Of course, after-images are unlike such things as trees and houses in certain obvious ways, and they are like pains in other ways, but the differences and likenesses are often exaggerated, and I think that some of Malcolm's arguments provide examples.

Malcolm claims that the concepts of numerical identity and numerical difference do not apply in the case of after-images. He thinks it is an absurdity to claim that two people see numerically the same after-image or that they see numerically distinct after-images. This is supposed to be a point of contrast between after-images and things like trees or houses, and a point of similarity with pain. We are invited to consider the following experiment. Suppose two persons gaze steadily at a piece of paper cut in the shape of a four-pointed star placed on a black background, and they then both turn their eyes to one and the same sheet of white paper. They see a grey star-shaped patch—the after-image. Suppose each traces the location of his after-image with his finger, and it turns out that they trace the very same path. Now we ask: Did the two persons see one and the same after-image or did each person see a distinct after-image? Malcolm holds that it does not make sense to say either, and his reason is that there is no further investigation we could carry out to decide the question.

I think the following little investigation would be relevant. We ask the two subjects in the experiment to shift their focus of attention slightly in opposite directions while keeping the after-image in view, and then to trace again the place where the after-image was. If they mark out different places, then they are obviously seeing their after-images lying in different directions, and it is inconceivable that a single star-shaped thing should lie in different directions at the same time. The investigation seems to me to show that the two subjects are seeing distinct after-images. One can imagine the circumstances turning out differently. Let our two subjects be called *A* and *B*; then it is conceivable that *A* should fail to keep the after-image in view except in so far as he shifts the focus of his attention in step with the shifts in *B*'s focus of attention. A puzzling case no doubt, and one might doubt if what *A* and *B* saw should now be called an after-image, despite the conditions being in every other respect as described in Malcolm's experiment. Here we must either (a) concede that it is conceivable that two people should see the very same after-image, or (b) hold that a single thing's being seen by two people is enough to discount it as an after-image. To concede (a) is to grant my point against Malcolm, and to hold (b) is to insist that after-images must be such that no two people can see each other's; each necessarily sees his own. According to the latter position we have no right to call what *A* and *B* see an after-image unless we are sure that *A* does not see the very star-shape which *B* sees, and vice versa. In other words, on the present view, there must be a star-shape which *B* sees, and a distinct star-shape which *A* sees—there must be two star-shapes before they could even count as after-images at all. Thus Malcolm is wrong whichever way we look at it.

---

[1] *Knowledge and Certainty* (Englewood Cliffs, N. J., 1963), pp. 73–95.

The case of pains seems to me to be very different. If, from the fact that John has a headache and Mary also has one we "conclude" that "at least two headaches exist," this would only be a bizarre way of saying that each of two people has a headache. That is precisely what Malcolm does say of after-images (*Knowledge and Certainty*, p. 77). But if my remarks above are correct, then there is room for a kind of investigation relevant to the question of after-image which is not possible with respect to pains. Thus, however different after-images are from countable physical objects, they are not as different as pains are.

II

Can we count parts, or multiple features, of an after-image? Malcolm concedes that we can, in a sense, do this, but he argues that the logic of counting is different here from what it is when we count things like ink-dots on a piece of paper. One of his reasons for saying this is that we cannot check on the count that we give in the case of an after-image. We cannot check, he argues, because there is no such thing as looking *at* the after-image at all, for it makes no sense to speak of looking away from it. The argument goes as follows.

1. If you can look away from something, then it can continue to exist when you no longer see it.
2. Your after-image cannot exist when you no longer see it.

Therefore,

3. You cannot look away from your after-image.

The first premiss is false. We can look away from something without entirely losing sight of it, so it is not out of the question that something whose being consists in being perceived should continue to exist as we look away from it, so long as we do not look so far away as to lose sight of it altogether. Of course, we cannot, as a matter of fact shift our attention away from our after-image even in this limited way, but we understand what it is to try to do so; we know what it is to fail to do so, and what it would be like to succeed. Thus it seems to make perfectly good sense to talk of looking away from an after-image. Even if we define an after-image in such a way as to exclude the possibility of looking away from it in this limited way, that would imply that one necessarily looks at an after-image, not that it makes no sense to say that one looks at it. However, such a definition would not do, for just as there are cases where one cannot look away from an after-image there are also cases where one cannot look directly at it either. If the original stimulus object was a bit to the left of one's focus of attention, then the subsequent after-image will stay stubbornly to the left of one's focus of attention no matter how one moves one's eye around. We know what it is to look askance at an after-image just as we know what it is to look directly at one. *A fortiori*, we know what it is to look at one.

There is yet another way in which one may look at, or fail to look at, one's after-image, and this is something one can, with some effort, control at will. One can focus one's eyes on an after-image that is directly in front of one's eyes and thereby see it very clearly, or focus one's eyes on something in the far distance that happens to be in one's direct line of vision, thereby taking scant notice of the after-image or perhaps not noticing it at all. The shift of focus here is comparable to the shift involved in looking at a spot on a window and then looking through the window into the far distance; a shift in the position of the eyes is involved in both cases. Just as one can choose to look at the spot on the window, or at what lies beyond the window, and change one's focus at will, so, with a little practice, one may change the focus of one's attention from an after-image to something in the distance, and back again. This means that there is such a thing as having a second look at the same after-image, which also means that we could take a second count of the features of the after-image, thereby checking on the first count. How do we know that we are refocusing attention on the same after-image? The answer is that we know enough about the way in which after-images are brought about to be able to decide whether we are having a second look at the same after-image, or seeing a new one. Not only can we distinguish between a bright, fresh after-image and a fading one; we can also usually tell where an after-image came from—we know which original stimulus object produced it. Thus we know quite well what it is to refocus attention on a fading after-image without having looked in the meantime at any stimulus object sufficient to produce that kind of after-image.

Malcolm admits that there is some sense in which we can be said to count the features of an after-image. He says that he could determine that his star-shaped after-image "has seven points by counting in succession from one to seven, pointing [his] finger at a different point of the after-image each time [he] utters a numeral" (*Knowledge and Certainty*, p. 80). Clearly, this can be done without shifting the direction of one's attention. It is a relatively simple task, given a small number of dots arranged into a small circle around a central dot, to fix one's attention on the central dot and count the number of dots around it by pointing at the dots as Malcolm describes. Where the number of dots is as low as two or three, it would hardly be a case of counting either, but of telling by looking. This can be done equally well whether it be a case of counting the points of a star-shaped after-image or a case of counting ink-dots on paper, but Malcolm would hold that the logic of counting is different in the two cases. His reason for saying that the logic of counting is different here is, as we have seen, that we cannot check on our counting in the case of the after-image as we could in the case of the ink-dots on paper; and we cannot check, he says, because we cannot be said to look at an after-image, or to look at it more carefully, or to have a second look at it. Against this I have argued that one can be said to look at an after-image, that one can be said to make sure it is in the direct line of vision as against being to the one side, and that there is something which can properly be called having a second look at the after-image.

There is also another form of checking that is available to us in the case of after-images. It will be generally agreed that the cause of our seeing an after-image is the stimulus object at which we have been previously staring, and that this stimulus-object determines the features of the after-image—its shape, color, and the number of its parts. Why, therefore, should we not check a man's report of his after-image by finding out what stimulus objects he had been previously looking at? Let us look again at Malcolm's experiment where a person looks at a four-pointed star and then transfers his gaze to a white sheet of paper to get the usual after-image. Suppose someone did this and then declared that as an after-image he saw a five-pointed star. Then we would be as surprised as we would be if he stared at the printed four-pointed star in the first place, and solemnly declared that he saw a five-pointed star there. In the latter case we would say that our subject was not telling the truth, or that there was something wrong with his eyesight, and we would say the same for the case of the after-image, for we know perfectly well what the perceiver *should* see in each case.

It is worth looking at these two cases a little more closely. When a person looks at a printed four-pointed star, and says he sees a five-pointed star, then we may say that he does not see the star as it is printed on the page, but in the case of seeing the after-image we have some reason for refusing to describe it as a case of failing to see the after-image as it is. There is, in the first case, a genuine four-pointed star printed on the page, a star which the percipient failed to see properly, but there would appear to be no genuine four-pointed after-image which the percipient failed to see properly in the second case. Thus, it is tempting to say that there is a check in the first case which is not available in the second case, and as soon as one says this, one is well on the way to accentuating the difference between the two cases in the way Malcolm does. This view of the matter is, however, mistaken.

If a person reports seeing a printed five-pointed star on a page, then indeed his report is correct only if there is a page with a five-pointed star on it. Once the percipient specifies the page and the star, we can say that if his report is correct, then that particular star should have five points. If not, there is something wrong with his report. In this case the percipient's report is such that, if it is to be accepted as correct, we require that what he reports seeing should be there in front of his eyes at the time of his seeing it, and its shape must be just as the percipient reports it. The possibility of our saying that the percipient's report is right, or is wrong, rests squarely on this requirement. However, it would be a big mistake to think that the possibility of judging a percipient's report for rightness or wrongness rests on precisely this sort of requirement in all cases. If someone reports seeing a star at a particular point in the night sky, then it could be one of those stars which the astronomers tell us is extinct, and this fact together with other facts about the speed of light, and the distance of stars, come into the requirement on which we base our judgment on the rightness or wrongness of the subject's report about seeing the star. We do not judge his report to be wrong simply on the grounds that the star is extinct. In other cases the requirement will be different again. For example, if the report is about seeing a mirror-reflection of a tomato on the window-sill, then we require that there be a tomato on the window-sill suitably related to the mirror, but we do not require that the mirror-image be an edible, squashable thing. Again, if someone who sees a scarf under artificial lighting conditions say "I see a scarf which looks quite black," we do not, in judging the correctness of his report, require that there should be a black scarf which he sees; it could turn out to be a dark navy-blue one. Thus there are requirements on the basis of which we judge a percipient's report to be right or wrong, but the requirements vary considerably as between different types of cases.

Now, to return to the case of the after-image, what we require here is that there should have been something at which the percipient gazed just before shifting his attention and seeing, as he claims, the after-image, and we require that this original stimulus-object should have certain features relevant to the after-image. We do not require that the after-image should have the same color as the stimulus-object but neither do we relax the requirement so much that the color of the stimulus-object is totally irrelevant. With regard to shape, we require that a five-pointed star-shaped after-image should be the result of looking at a five-pointed star-shaped original stimulus object. This requirement arises from our grasp of the concept of an after-image just as certain other, different requirements spring from our concept of a mirror-image. In each case they are requirements which provide a basis for a judgment of the rightness or wrongness of the relevant perceptual reports.

All that is necessary for the corrigibility of a person's report is that it should be possible for others to tell, without asking the person himself, whether a particular necessary condition of the truth of that report holds. It is obvious that the claim to see a book on the table is corrigible by this standard, for the truth of the claim requires that there be an independently existing book which we can independently inspect, but the claim to see an after-image is also corrigible by a comparable standard. Granted, the truth of this claim does not require that there be an independently existing after-image which we can independently inspect, but there are certain other conditions which must hold if the claim is true, and it is possible to tell whether they hold without asking the person who makes the claim. There must at least have been an object of a specific kind at which the percipient was looking prior to seeing the after-image, and we can check whether it was so independently of what the percipient has to say. For instance, if a person who has been asleep for twelve hours claims that the very first thing he saw on opening his eyes was an after-image, then we know that his claim is false. Whatever a man sees under such conditions, it cannot be an after-image.

Consistently with this, it makes sense to talk of a person having a distorted view of an after-image in much the same way as it makes sense to talk of a person having a distorted view of a tangible, visible object. Imagine an experimental situation in which electrodes are inserted so as to make it possible electrically to stimulate a certain area of the brain. Suppose the subject looks at a circle which appears circular to him until the electrical stimulation is applied when the circle suddenly appears oval. Are we to say that the subject stops seeing the circle as soon as the electrical stimulation is applied, and that he sees something else instead? Obviously, we should say that he still sees the circle, only that he gets a distorted view of it. Similarly, if a person looked at a yellow circle and a similar electrical stimulation was applied as he transferred his gaze to a blank grey page, and he reported seeing an oval purple after-image, then we should say that he was getting a distorted view of the after-image which he should see. The circumstances are too much like the normal circumstance of seeing the after-image for it to make sense to say that he was not seeing the after-image at all, or even to say that he was having a different after-image.

It is instructive at this point to contrast the claim to see an after-image with a claim to see, for example, what is simply described as a green circle. The green circle could turn out to be a metal or wooden figure; it could be a collection of green dots placed among some pink ones to test for color blindness; it could be the after-image resulting from staring at a light filament, and it could also simply be a green circle seen by the percipient and nothing more. Thus there are no conditions necessary for the truth of the claim to see a green circle such that those conditions could be known to hold by someone other than the claimant, and known without asking him. We could never be in a position to tell a person who claims to see a green circle that he cannot be seeing one, or that he must be getting a distorted view of something or other. Of course it is a different story if we have a description like "a green circle chalked on the blackboard," for a claim to see something so described is clearly open to challenge by an independent observer. The point is that if one merely claims to see a green circle then one commits oneself to nothing that could be challenged by another on independent grounds. Such a claim is indeed incorrigible, but it contrasts sharply with the claim to see an after-image. If someone claims to see an after-image, then we could be in a position to say that he did not see one, or that he did not see the after-image which he claimed to see, or that he got a distorted view of it.

The difference between these two claims lies in the implications of the descriptions. My description of the object of my perception as "an after-image" carries the implication that certain conditions hold, and they are conditions which could be known to hold by others independently. Such implications do not follow from the description of what one sees as a green circle. The point is of some importance since there could be other, and for our purposes irrelevant, grounds on which a percipient could be told that he is not seeing what he should see, or is seeing what he should not see. Suppose a psychologist puts a number of subjects in a dark room and has a set-up for electrically stimulating the visual area of his subject's brains in a specific way. Everytime this is done they see a red circle. This is tried out with hundreds of different subjects until one day one of these subjects, under apparently identical conditions, reports seeing a green circle. The psychologist might well say that the percipient is not seeing what he should see, but this is no challenge to the validity of the percipient's description of what he sees as a green circle. It is one thing to say, on the basis of some psychological knowledge or theory, that a person is seeing what he should not see, but quite another to say that his report is incorrect. I fail to see what grounds there could be for saying that a person's report, that he sees a green circle, is incorrect (assuming, of course, that the person knows and means what he is saying). It is in this respect that a claim about seeing an after-image contrasts sharply with a claim to see a green circle, and it is in this sense that the former is corrigible and the latter not so.

I conclude that Malcolm is mistaken in thinking that there is no check relevant to reports about the number of parts of an after-image and that we should not accept his suggestion that the logic of counting parts of after-images is different from the logic of counting things like the panes of a window, or the points of a printed star-shape on a page. Taking this point together with the point argued in the previous section we cannot but conclude that Malcolm has exaggerated the difference between seeing after-images and seeing things like tables and chairs. There are differences undoubtedly, a variety of them; but the similarities are rather interesting too.

III

A different attempt to throw light on the problem of perception is to be found in Miss Anscombe's article on the intentionality of sensation.[2] She distinguishes between intentional and material objects of perception, and her very brief references to after-images in this context implies that they are intentional, not material, objects of perception. An after-image is contrasted with a real spot (a stain) in such a way as to imply that the latter, but not the former, is a material object of seeing. She also says that if we take the Platonic dictum "He who sees must see something" to mean that "there must be something there which is seen" then we legislate against all except the material use of "see." The implication here is that the material object of seeing is something which is there, whereas an intentional object such as an after-image is not something which is there. A material object is held to be something that "really—physically—exists," but not so an intentional object such as, I think, she takes an after-image to be.

As against this, one would very much like to say that after-images are really there in front of one's eyes. Contrast an after-image with, say, a pain, or a thought, or a taste which is in no sense there before your eyes. These could not conceivably interfere with your view of a white door which is in front of your eyes, whereas an after-image can do so. There would appear to be *some* sense in which an after-image is there in front of you, but there is *no* sense in which a pain or a thought is there in front of you. Again, we may contrast after-images with centaurs. Centaurs could conceivably interfere with your view of other things that happen to be in front of you, but none ever do, for they do not exist. There are after-images, and they are more or less in front of your eyes.

I think we may go further than this, and say that the after-image is at a certain distance from the perceiver. Suppose you stare at the glowing red bars of an electric fire for half a minute or so, and then look at a white wall so that the usual green after-image appears on the wall. Then approach the wall keeping the after-image carefully in view, and you will find that the after-image diminishes in size relative to the wall as you approach it. As is well known, this is precisely what happens if anything that you look at as you approach the wall keeps its distance from you as you move forward. The after-image keeps its distance from the perceiver relative to anything which lies around, and it takes little reflection to realize that the distance kept is the distance between the perceiver and the original stimulus object which was the cause of the after-image. Just as the original object determines the shape, color, and number of parts of the after-image, so it determines the distance as well.

---

[2] "The Intentionality of Sensation: a Grammatical Feature," *Analytical Philosophy: Second Series* (Oxford, 1965), pp. 158–180.

Admittedly, the after-image does not have a volume, and it does not occupy a place such that other things have to be shifted to make room for it, but there is no denying that it has a position, and that it lies at a certain distance in a certain direction. It is unlike some other things that are at a distance from us in that the perceiver cannot change the distance between himself and the after-image; he cannot get nearer to it or move away from it, though he can change its direction at will by turning his eyes in whichever direction he wants. Furthermore, if I approach the wall keeping the after-image in view there can be stages where the after-image is between me and the wall, then a stage when it is the same distance from me as the wall, and then as I get still closer to the wall, it will be farther from me than the wall, yet without getting behind the wall in the sense of being hidden from view by it. The after-image certainly behaves oddly by contrast with the chairs, books, and so forth that lie around, but none of this implies that the after-image does not have a position, or does not lie at a certain distance from the perceiver, and in a certain direction.

IV

I have argued that after-images really do exist and are there in front of one's eyes (or possibly a little to the side, depending on the way we saw the original stimulus-object). It is no objection to this to say that the after-image cannot be touched, or does not have a volume, any more than saying this would be an objection to the claim that a rainbow, a shadow, the sky, or some oxygen is really out there. If we know what sort of thing each of these is, then we know what to expect of them, and we need not be misled by the claim that they really exist out there.

Yet there is, surely, one glaring difference between the after-image and all the other examples that I have just mustered. They are things that can play the role of causal conditions in the affairs of the physical world in a way that after-images apparently cannot do. Admittedly, my examples cannot all play the same kind of causal role. Oxygen can help to keep a fire going, whereas none of my other examples could do that, but even shadows, rainbows, and the sky can play a causal role in producing a certain effect on a photographic print. An after-image cannot even do that. One can also say of a shadow, or the sky, or a rainbow that it is the cause of our seeing it, whereas an after-image is not the cause of our seeing it. One might, on this ground, say that an after-image, even if it really exists and is there in front of one, is not a physical reality. We now seem to have a criterion which puts after-images where Miss Anscombe wants to put them, namely, on the other side of the fence from what she regards as things which really, physically exist.[3]

This criterion should also be attractive to Malcolm, though it is different from the one which he considers. He distinguishes between objects of direct perception, and other objects of perception, on the ground of the incorrigibility of the perceiver's report of the former, as contrasted with the corrigibility of his reports of the latter. But according to the arguments that have been produced above, the percipient's reports on his after-image are corrigible, and that implies that Malcolm's criterion should have led him to put after-images on the other side of the fence from where he does put them; Malcolm's own criterion, properly applied, would leave after-images in company with trees, houses, stones, and so forth. The causal criterion which has just been mentioned in the last paragraph, however, puts after-images where Malcolm wants to put them, in a category apart from trees, houses, stones, and the like.

Still, I doubt if this causal criterion, and the distinction effected by it, should be elevated above others as though some special importance attached to it. Malcolm's own criterion of corrigibility is no less valid if properly applied. As I have already argued, reports about after-images are indeed corrigible, whereas reports about seeing a green circle are not. The interesting thing which the case of after-images shows is that these two criteria do not coincide. After-images go together with incorrigibles by the application of the causal criterion, but they go together with material, physical objects by the application of the incorrigibility criterion. What it comes to is that we have a threefold classification to say the least—a classification which puts incorrigibles together on the one side, and material, physical objects together on the other side, but which also leaves a place in between for after-images.

---

[3] For a possible account of the causal connection see R. Chisholm in *Perceiving* (Cornell, 1957), ch. 10.

Even apart from the reference to corrigibility there is a further difference between seeing a green circle (which is *simply* a green circle—not a green circle drawn on a piece of paper, or even an after-image) and seeing an after-image. Though both are alike in not being the cause of our seeing them, they are different in that an after-image is an image of the cause of our seeing it, whereas a green circle is not. A green circle, the report of seeing which is incorrigible, may be caused by anything, including such things as taking a drug, or having one's brain electrically stimulated, and the description of what is seen need not be true of the cause of one's seeing it. It is not so with the after-image; an after-image must be an image of that which caused one's seeing it for it to count as an after-image at all. Thus, although seeing an after-image is different from seeing a chair, it is also different from seeing a green circle, and there is no obvious reason why one should make more of the one difference than the other. If we want to distinguish between objects of perception, then the case of after-images seems to indicate that it is a trichotomy, rather than a dichotomy, that is needed.

All this has obvious implications for the application of the notion of an intentional object of perception. If an intentional object is taken to be something which we cannot be mistaken about, then we would not (as Miss Anscombe does) count an after-image as an intentional object of perception. But if we insist that after-images are intentional objects, then we are left with a division within the category of intentional objects between those which are corrigible and those which are not, and between those which are the image of the cause of one's seeing them, and those which are not. The difference between the two sides of this division would appear to be no less worthy of attention than the difference between the categories of intentional and material objects of perception.

V

There are other so-called intentional objects of seeing which deserve to be distinguished sharply from after-images. Consider the example of a mirage (from M. Luckiesh) which Miss Anscombe quotes—"An approaching pedestrian may have no feet (they are replaced by a bit of sky)."[4] We may say, according to Miss Anscombe, that the material object of seeing in this situation is a pedestrian with feet, but she would insist that it is also true of what is seen, in some sense, that it is a pedestrian without feet. Now we may well ask, in what sense is this true of what is seen? It may be argued that it is true at least in the sense that the pedestrian looks as though he did not have any feet (though he does have them) but as soon as this is said the difference between this case and the case of the after-image comes clearly into view, for there is nothing other than the after-image of which it is true that it looks green and circular (while not being green and circular) when one sees a green circular after-image. It will not do to say that a circular part of a wall looks green, for one can still see the after-image even with one's eyes closed.

A pure mirage is, in some ways, more like an after-image. If one sees what seems to be a stretch of water in the distance in a sandy desert, then it is hardly true of anything other than the mirage that it, while not being a stretch of water, looks like a stretch of water. Just think how different it would be if there was a huge mirror out there in the desert, and which looked exactly like a stretch of water; seeing such a thing would not be a case of seeing a mirage at all. A mirage compares with an after-image in another way, and in this respect both contrast with what is, say, a mere green circle. To describe what one sees as a mirage is to imply something as to the causal conditions of one's seeing it, which is a point that I have already argued for in the case of the after-image. The causal conditions implied are, naturally, different in the two cases; a mirage is due to atmospheric conditions. But both the mirage and the after-image contrast with the mere green circle in that to call what one sees a green circle is not to imply anything about the cause of one's seeing it. Yet there are differences here too. For example, the mirage does not have to resemble that which is the cause of one's seeing it for it to count as a mirage, whereas it is otherwise with after-images.

I have been considering objects of seeing which could be classified as merely intentional, and as non-material, by the application of the causal criterion. These have turned out to be different in various interesting ways, and I have suggested that these differences are no less important than the difference shown by the application of the causal criterion. This point is reinforced if we look again at some of the things that seem to fall on the other side of the fence, things that would appear to rank as material objects by the application of the causal criterion. The particular examples I have in mind are shadows, rainbows, and the sky. It seems to be quite in order to say that these produce certain effects on photographic film, and that they are also the cause of our seeing them. Clearly the causal account to be given will be rather different from the kind which Chisholm outlines for seeing things like tables and chairs, principally because shadows, rainbows, and the sky cannot be said to have surfaces, but this does not seem to be a sufficient reason for saying that they cannot be regarded as causes at all. However, these examples compare with after-images, and contrast with things like houses and trees, on other grounds. They do not occupy a certain volume of space. The sky and rainbows can be located even less precisely than after-images, and they are all equally diaphanous. None of them can be touched or handled, as books and pencils can. The list of comparisons and contrasts could be extended. The upshot seems to be that the distinction effected by the causal criterion should not be over-rated at the expense of neglecting other distinctions.

---

[4] *Analytical Philosophy*, *op. cit.*, p. 170.

VI

We have seen something of the futility of invoking the notion of intentionality to effect a dichotomous division of objects of seeing. I now want to consider one last example, not only to show the extent of the muddle we can get into if we try to distinguish between objects of perception in terms of intentional and material objects, but also to throw a hint as to the real role of intentionality so far as perception is concerned.

We can easily imagine circumstances in which someone thought for a moment, with some reason, that he was seeing an after-image against a wall, though it was not really an after-image at all; it was a discoloration of a part of the surface of the wall. Did the man see the discoloration on the wall? We may feel like saying "Yes" and "No" to the question, and someone who believes in the intentionality of seeing might explain it in this way. He suggests that the answer "Yes" is right in that the material object of seeing was the discoloration on the wall, and that the answer "No" is also right in that the discoloration was not the intentional object of seeing. He might add that the intentional object of seeing in this example is an after-image. The suggestion might be put a little differently. It might be said that in the material sense of the word "see" the discoloration on the wall was seen, but that in the intentional sense of the word "see," it was an after-image that was seen. Again, it might be put this way—the words "a discoloration on the wall" give the material object of seeing, whereas the words "an after-image" give the intentional object of seeing.

It is easy to see how misguided this account is if we consider circumstances which are the reverse of those just described, that is to say, circumstances in which a man honestly says that he sees a discoloration on a wall, when in fact it is an after-image that he sees. How could the above account be plausibly adopted for this case? Should we say that the after-image was the material object of his seeing, and a discoloration on the wall was the intentional object? Would it be right to say that in the material sense of the word "see" an after-image was seen, but that in the intentional sense a discoloration was seen? Will it do to say that the words "an after-image" give the material object of his seeing and that the words "a discoloration on the wall" give the intentional object of his seeing? Clearly, the suggestion is altogether wrongheaded.

Suppose we consult the perceiver himself after he has realized his mistake. Once he realizes that he is seeing an after-image, he is no longer prepared to say that he saw a discoloration on the wall in any sense of the word "see," nor that there was some non-material seeing involved. He will put us right by saying that the truth of the matter is that he *thought* (believed) that what he saw was a discoloration on the wall, and that in thinking this he was quite mistaken. There seems to me no reason why we should not go along with him in this, and if we wish to use the word "intentional" at all in this connection, then we might use it for referring to the element of thought (belief) which so often accompanies seeing.

This hint raises a topic which deserves a fuller discussion than can be given here, but if it is along the right lines, then we may at least note that the only case of seeing which could properly be called intentional would be the case where the verb takes a propositional object. It is one thing to see an after-image, and a different thing to see (in a second sense) that what I see (in the first sense) is an after-image. Seeing, in this second sense is what I suggest might be called intentional. If this is right, then it is out of the question that we should distinguish between after-images, on the one hand, and tangible objects on the other, in terms of intentional and material objects of perception. On the view now mooted, the object of seeing in the intentional sense will be propositional, and seeing in this sense is not something we could do with our eyes, but with our minds.[5] That this is so becomes obvious in the case where a mistake is discovered. In that case we withdrew the claim to have seen that, . . . and substitute the claim merely to have thought that . . . (for "see" carries an implication of success).

As regards seeing (as distinct from having thoughts as to what one sees), this study of after-images indicates that there is a great variety of differences and likenesses between a variety of objects of seeing—differences and likenesses which are easily missed if we persist in looking for a dichotomous division of objects of perception. I would suggest that paying attention to the details of the variety of differences and likenesses can safely, and profitably, be allowed to displace the broad theoretical distinctions so often invoked, especially since such theoretical distinctions are apt to generate philosophical puzzles of their own, puzzles which dissolve again as the details are brought into view.[6]

University College of Wales,
Aberystwyth

Received February 26, 1971

---

[5] My suggestion here is in line with the arguments produced by A. W. Collins in "The Epistemological Status of the Concept of Perception," in *The Philosophical Review*, vol. 76 (1967), pp. 436–459.

[6] I am grateful to my colleague, Mr. Ian Tipton, and staff and students at the University of Keele, and the University College of Swansea for discussion of earlier drafts.

American Philosophical Quarterly
Volume 9, Number 2, April 1972

IV. IS RELIGION A "FORM OF LIFE"?

PATRICK SHERRY

I. Introduction

IN his review of Ludwig Wittgenstein's *Philosophical Investigations* Norman Malcolm said that one could hardly place too much stress on the importance of the notion of a "form of life" in Wittgenstein's thought.[1] Later on he applied this concept to religion in his Memoir of Wittgenstein,[2] and especially in his article on St. Anselm's Ontological Arguments. In the latter he said that he suspected that the argument could only be thoroughly understood by someone who understood from the inside that "form of life" which gives rise to the idea of an infinitely great being and who therefore had some inclination to partake in that religious form of life.[3] Wittgenstein himself, as far as I know, never described religion as a form of life, but he came very close to so doing when he asked in his lectures on Religious Belief "Why shouldn't one form of life culminate in an utterance of belief in a Last Judgement?"[4]

Unfortunately Malcolm did not develop his identification of religion with a "form of life" any further. But it is quite clear that he had an axe to grind here, and that he is not merely repeating the familiar point of apologists that religion is a way of life and not just a body of doctrines. For he also applied the closely associated notion of a "language-game" to religion, attempting to establish the following two points:

(i) religion is "given" or basic, so that it need not be justified and indeed cannot be: "this language-game is played."
(ii) its concepts are in order as they are.

A third point, suggested by Malcolm but developed more fully by Peter Winch, Paul Holmer, and D. Z. Phillips, is that religion can only be properly understood from the inside, because it is a complex of certain human activities, responses, and relationships. We can only understand the "grammar" of religious beliefs by understanding this complex in which they have their life, and to get this understanding we need a "feeling for the game."[5]

Fortunately we can consider whether religion is a form of life without having to agree with the points made by these writers. My purpose in this paper is to see in what sense religion can be described as a "form of life," and what light such a description throws on the philosophy of religion. My conclusion will be that religion *includes* forms of life, rather than that it *is* a form of life, and that the writers I have mentioned have misrepresented the concept and used it in a question-begging way. I think that their approach is perhaps fruitful in showing us what kind of an activity religion is and what role it plays in people's lives, but that it needs supplementing when we come to consider the character of religious experience and the nature and evolution of religious doctrine. In so supplementing it we shall need to separate two strands in Wittgenstein's thought, which I shall call the Relativist/Conventionalist and the Pragmatic/Empiricist, and to follow the latter.

First of all, however, I want to examine the concept of a "form of life" carefully and to see what role it plays in Wittgenstein's thought generally. I think that this procedure is necessary because of the common misunderstanding of the concept. Dallas High, for example, speaks of the "... claims of the natural scientific point of view which roughly coincide with modernity or the context of the form of life in which we live in the twentieth century."[6] Whatever that may mean, it is far from Wittgenstein's usage.

II. Wittgenstein on "Forms of Life"

Wittgenstein only uses the expression "form of life"[7] five times in the *Philosophical Investigations*[8] but we can supplement these references with passages from his other works, and also with allied concepts like "activity," "custom," "natural history," and "the common behavior of mankind." He introduces the notion when he says that "to imagine a language means to imagine a form of life."[9] In his Lecture on Aesthetics he had said that the main mistake made by Moore and most contemporary philosophers was to look at language merely as a form of words, and not at the use made of them: "Language is a characteristic part of a large group of activities—talking, writing, travelling on a bus, meeting a man, etc."[10]

---

[1] *The Philosophical Review*, vol. 63 (1954), p. 549.
[2] *Ludwig Wittgenstein: A Memoir* (London, 1958), p. 72.
[3] *The Philosophical Review*, vol. 69 (1960), p. 62.
[4] *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*, Cyril Barrett (ed.), (Oxford, 1966), p. 58.
[5] D. Z. Phillips, *The Concept of Prayer* (London, 1965), p. 83.
[6] D. M. High, *Language, Persons and Belief* (New York, 1967), p. 174.

Wittgenstein gives only one specific example of a "form of life" in the *Philosophical Investigations*: on p. 174 he says

> Can only those hope who can talk? Only those who have mastered the use of a language. That is to say, the phenomena of hope are modes of this complicated form of life.

I take it (following J. F. M. Hunter, *op. cit.*, p. 240) that the form of life being described here is using language rather than just "hoping."[11] This interpretation is confirmed by P.I. 19, already quoted, and by P.I. 23, where Wittgenstein says that speaking a language is "part of an activity, or of a form of life." Presumably we may regard the "language-games" which he lists there, e.g., giving or obeying orders, making up stories, joking, asking, thanking, praying, cursing, and greeting, as forms of life, too. Wittgenstein gives this list to make the point that language can be used in all sorts of ways besides that of describing things.

A more puzzling passage is this: " 'So you are saying that human agreement decides what is true and what is false?'—It is what human beings *say* that is true and false, and they agree in the *language* they use. This is not agreement in opinions but in forms of life" (P.I. 241). The next section explains this to some extent: if language is to be a means of communication people must agree in their judgments as well as in their definitions of words. "Measuring" depends on people's getting the same results, just as using color concepts presuppose that there is general agreement about how things look (P.I. 242 and p. 226).[12]

Another example, given in his notes "On Certainty," is feeling certain, or rather having a "comfortable certainty." This, says Wittgenstein, isn't something akin to hastiness or superficiality, but is a form of life: "But that means I want to conceive it as something that lies beyond being justified or unjustified; as it were, as something animal" (357–9).

The above remark on justification ties in very well with the one remaining reference in P.I., where Wittgenstein says "What has to be accepted, the given, is—so one could say—forms of life" (p. 226),[13] and with some other remarks about justification, e.g., "If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: 'this is simply what I do' " (P.I. 217); or "Our mistake is to look for an explanation where we ought to look at what happens as a 'proto-phenomenon.' That is where we ought to have said '*this language-game is played*' " (P.I. 654).[14]

The refrain "This is what I [or we] do" is one frequently encountered in Wittgenstein's later philosophy. Usually it is introduced to counter what he regards as a mistaken hankering after justification (cf. R.F.M. II. 74). Wittgenstein insists that explanation must at some point give way to description (P.I. 109) and that it is sometimes silly to ask for reasons for things: "Do we live because it is practical to live?" (R.F.M. V. 14). "Does man think, then, because he has found that thinking pays . . . Does he bring his children up because he has found it pays?" (P.I. 467).[15] No, for sometimes we must just say "such is human life."[16] Giving grounds comes to an end some time, but "the end is not an unfounded presupposition: it is an ungrounded way of acting" (*On Certainty*, 110). This brings us to bedrock: "The limits of empiricism are not assumptions unguaranteed or intuitively known to be correct: they are ways in which we make comparisons and in which we act" (R.F.M. V. 18).

---

[7] See C. A. Van Peursen, *Ludwig Wittgenstein: An Introduction to his Philosophy*, tr. by R. Ambler (London, 1969), pp. 108–109, for some suggestions about Wittgenstein's derivation of the term.

[8] Cf. J. F. M. Hunter, " 'Forms of Life' in Wittgenstein's Philosophical Investigations," *American Philosophical Quarterly*, vol. 5 (1968), pp. 233–243.

[9] *Philosophical Investigations*, Pt. I. §19 (This will henceforth be referred to as P.I.; references to the sections of Part I will be given simply in figures, and references to Part II will be given by page number.)

[10] *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*, ed. by Cyril Barrett (Oxford, 1966), p. 2.

[11] Though I think that Wittgenstein also regards hope as a form of life: elsewhere (P.I. 583) he describes it as a "phenomenon of human life."

[12] He makes a similar point in the *Remarks on the Foundations of Mathematics* (henceforth cited as "R.F.M."): "Calculating would lose its point if *confusion* supervened. Just as the use of the words 'green' and 'blue' would lose its point" (R.F.M. II. 75; cf. also *Zettel*, §§351, 430).

[13] A text appealed to by D. Z. Phillips, *op. cit.*, p. 27.

[14] Likewise a text appealed to by Malcolm, "Anselm's Ontological Arguments," p. 56.

Sometimes he speaks in a similar vein of "natural history": "commanding, questioning, recounting, chatting, are as much a part of our natural history as walking, eating, drinking, playing" (P.I. 25; cf. 415). This even applies to mathematics, for measuring, weighing, and costing are procedures of human behavior, and going through mathematical proofs and accepting their results is "simply what we *do*: this is use and custom among us, or a fact of our natural history" (R.F.M. I. 63; cf. 146).[17] He constantly stresses the practical applications of mathematics, e.g., when he says "The technique of calculation is part of the technique of house-building" (R.F.M. I. 142), and insists that pure mathematics isn't merely an isolated abstract intellectual discipline. Even something like the concept of the rule for the formation of an infinite decimal "is a concept connected with a rigidly determined *activity* in human life" (V. 35). Hence he prefers to describe mathematics as an "activity" (P.I. p. 227) or an "anthropological phenomenon" (R.F.M. V. 26).

Other concepts of Wittgenstein which need to be reckoned with here are "The common behaviour of mankind" (P.I. 206), "culture" (*Zettel*, 164), "customs" and "institutions." Examples which he gives of the last two are obeying rules, making reports, giving orders, playing chess (P.I. 199), using money (P.I. 584) and punishing criminals (*Lectures and Conversations*, p. 50). He points out that many of our psychological concepts depend on them, e.g., "An intention is embedded in its situation, in human customs and institutions" (P.I. 337). This is to be associated with his frequent insistence that we can only understand the meaning of concepts by studying the surroundings or circumstances of their use: "Could someone have a feeling of ardent love or hope for the space of one second—*no matter what* preceded or followed this second?—What is happening now has significance—in these surroundings. The surroundings give it importance. . . . (A smiling mouth *smiles* only in a human face)" (P.I. 583).[18]

So much by way of exposition. We are now in a position to see what a form of life is, and why the concept is so important in Wittgenstein's later philosophy. By "forms of life" he means basic human activities and responses like hoping, feeling certain, measuring, giving orders, asking questions, and greeting people, and indeed using language generally. He regards these things, together with customs, institutions, etc., as the fundamental "given" facts from which philosophy must start.[19] This forms a marked contrast with Carnap, with Russell in his Logical Atomism Period, and with other philosophers who regarded simple sense-perceptions (and propositions describing them) as the starting-point for philosophy. Of course, Wittgenstein rejects the view that reality has simple constituent parts to which everything can be reduced, on independent grounds (see P.I. 47, where he argues against Russell's Reductionism, pointing out that words such as "simple" and "composite" have several different uses). In so far as anything is basic for the later Wittgenstein, it isn't sense-data or atomic propositions, but forms of life—or simply "life."

III. RELIGION AND FORMS OF LIFE

Clearly this is very important for religion. Russell's Atomism and Carnap's reductive epistemology were both inhospitable to religion, because it is impossible to analyze most religious doctrines into simple propositions about sense percepts. But Wittgenstein seems to offer a much more hopeful starting-point, and one, moreover, which is in harmony with the insistence of many contemporary theologians that religious belief is a commitment or a way of life, rather than the acceptance of a series of propositions.

All the same, *is* religion a form of life? Surely my exposition suggests that it is incorrect to label it so; rather, religion includes several forms of life (in the plural) along with some of the other categories which Wittgenstein mentioned, for example, "activities," "language-games," "customs," and "institutions." Hence Malcolm and Phillips are wrong to speak of religion as a form of life (in the singular).

---

[15] In 470 he admits that we do *sometimes* think because it has been found to pay.

[16] "Bemerkungen über Frazer's 'Golden Bough'," *Synthese*, vol. 17 (1967), p. 236.

[17] He heads off some possible misunderstandings when he points out that although calculating is part of a technique which is a fact of natural history, of course its rules aren't propositions of natural history (R.F.M. V. 14–15).

[18] Cp. *Zettel*, *op. cit.*, §173: "Only in the stream of thought and life do words have meaning."

[19] Hence Dallas High is correct in saying that "The phrase 'form of life' functions as a logically primitive concept." (*Op. cit.*, p. 102).

Religious forms of life would include hoping and feeling certain: for if these are forms of life, as Wittgenstein says they are, then presumably their religious embodiments, i.e., the theological virtues of Faith and Hope, are so too. Other possible examples which spring to mind are loving, meditating, repenting, and forgiving. But the form of life of using a language is the one which has the most import here: if praying, cursing, telling stories, commanding, and obeying are language-games, as Wittgenstein suggests (P.I. 23), then this is obviously applicable to religion, since the first two examples are specifically religious, while the others can all occur in a religious context. Moreover, we might adduce confessing, absolving, reciting litanies, singing hymns, and reciting stories or parables as language-games, too. Wittgenstein's other categories, of "customs," "institutions," and "activities" are also relevant here: genuflecting, fasting, sacrificing, kissing relics, taking part in sacraments and rituals, and going on pilgrimages perhaps spring to mind.

Things like reciting creeds, preaching sermons, and engaging in apologetical arguments are perhaps more difficult to classify, and raise serious problems concerning evidence and truth. It is also arguable that "worshipping" is a form of life, but in fact this probably boils down to a number of things already mentioned, especially praying, loving, and taking part in ceremonies.

IV. Forms of Life and Justification

The point which I have raised about religion not being a form of life, but a collection of forms of life, language-games, and so on is important because it disposes of certain arguments put forward by Norman Malcolm and D. Z. Phillips about the illegitimacy of trying to justify religious belief. These arguments depend to a large extent on the claim that religion is a form of life or language-game. D. Z. Phillips, for example, says that when people lose their faith in God "what has become meaningless is not some feature of a form of life, but a form of life as such."[20] He maintains that such a form of life is, like all forms of life, beyond justification: philosophers who ask for one are like those who ask for a justification of science, whom Wittgenstein compared to the ancients who posited Atlas as a support for the world.[21] He does, of course, allow that religious believers may justify particular articles of the faith, since here there is an agreed-upon frame of reference. But he refuses to countenance any general justification of religious belief: ". . . there is no question of a *general* justification of religious belief, of giving religion a 'sound foundation'."[22] Norman Malcolm takes a similar line (which has been labeled by Kai Nielsen as "Wittgensteinian Fideism,"[23] arguing that although we can adduce grounds for particular doctrines, e.g., the Apostolic Succession or the Holy Spirit, we cannot do so in the case of God's existence, because we have no idea what would constitute good grounds here: "do we know what a good reason would look like. . . ?"[24] Phillips compares the question of the reality of God to that of the reality of the physical world, i.e., it is a presupposition of any enquiry rather than a conclusion.[25]

There is a lot of confusion in all this. In the first place, I doubt whether anyone does ever try to justify "religion" as such (as opposed to a particular religion). Missionaries and apologists are always concerned to justify their own particular religious beliefs. No doubt (as the context shows) Phillips has in mind philosophical theologians like Crombie and Mitchell who write about "religious language." But even they are not trying to justify religion as such, but to show that the metaphysical strand of Christianity and other theistic religions is meaningful. Similarly, philosophers who try to produce proofs for the existence of God cannot be ruled out of account simply because belief in God is a presupposition of religious belief: even many philosophers who consider that all so-called proofs for the existence of God are in fact invalid would disagree with Malcolm's contention that we do not know what it would be to produce arguments for His existence.

---

[20] "From World to God?", *Proceedings of the Aristotelian Society*, Supplementary Volume XLI (1967), p. 141.

[21] "Religion and Epistemology—some Contemporary Confusions," *Australasian Journal of Philosophy*, vol. 44 (1966), pp. 316–317.

[22] *The Concept of Prayer*, *op. cit.*, p. 27. Phillips cites P.I. p. 226 here, as a support for his conclusion.

[23] *Philosophy*, vol. 42 (1967), p. 191.

[24] In John Hick (ed.), *Faith and the Philosophers* (London, 1964), p. 108.

[25] *Religion and Understanding* (Oxford, 1967), p. 67.

As regards the question of justifying forms of life, Phillips' argument is vitiated by his regarding religion as a form of life, and not as I have urged throughout this paper, a collection of forms of life, language-games, institutions, etc. But let us put the best complexion on his argument and rephrase it by arguing that forms of life and language-games like hoping, praying, worshipping, feeling certain about things, and telling stories cannot and need not be justified. But this still won't do. Again, no one ever does try to justify hope as such, but they justify their own particular hopes. A Christian, for example, hopes for the forgiveness of his sins and for the resurrection of the dead. If a sceptic questioned these beliefs, it would not be to the point to urge that "hope" is a given form of behavior. Rather, the Christian would try to provide some justification for his beliefs by appealing to Christ's promises and to the fact of the Resurrection (I am not claiming that any such justification is a *simple* matter). Similarly, he justifies his worship of God by arguing that God is more worthy of worship than any alternative. When Wittgenstein said that forms of life are "given" and beyond justification or criticism, he meant that it is absurd to ask for a justification of hoping and so on in general, since "this is simply what we do." But this doesn't mean that we can't ask for a justification in *particular cases*, e.g., by asking whether a hope or an expectation is likely to be fulfilled. Likewise, when Wittgenstein said "this language-game is played," he meant, among other things, that it is simply a fact of nature that people tell jokes, give or obey orders, make up stories, and report on past events. It would indeed be silly to ask for a justification of such practices; but *individual cases* can be appraised in various ways—jokes can be funny or crude, orders pointless or immoral, and reports of events untrue or graphic.

The distinction which I have just drawn between forms of life as wholes, and particular exemplifications of them, is similar to the one which Phillips draws between religious beliefs in general and particular religious beliefs, so he might perhaps be expected to accept my contention that we can appraise *particular* exemplifications of forms of life—particular hopes, expectations, loves, etc. But it seems that he will not grant us even this: when discussing Wisdom's work in the philosophy of religion he argues that Wisdom couldn't really refute people who thought that Nero was the incarnation of God or who worshipped cats.[26] Well, it would depend on whom he was arguing against. But surely Christianity conquered ancient Rome precisely because Christ seemed a more plausible incarnation of God then Nero and its God seemed more worthy of worship than cats or idols? Moreover the early Fathers, unlike Phillips, were willing to argue their case and to deal with particular religious issues.

One might well want to press the attack on Phillips even further at this point by arguing that there are *some* activities, responses, forms of life, and language-games which can be appraised as wholes. We do, for example, think it appropriate to rejoice at good things and to be resigned to suffering (not vice-versa). We do criticize those who fail to show pity or affection, and consider that people should give thanks or show forgiveness. It is difficult to say how far we can justify ourselves here and to see what we should say to a race of men who never forgave, pitied, loved, or gave thanks. It would, of course, depend on why this was so: if they were constitutionally incapable of these responses we might feel that they could not be classified as human beings. But otherwise we would surely at least try to teach them to be loving, forgiving, and so on.

Phillips contends that there could be no general justification of such practices. Of course a Christian could justify the practice of forgiveness, for example, by appealing to Christ's teaching on this and on the love of our neighbor and also to his own example on the Cross, but Phillips considers that these criteria are themselves unjustifiable, for what would it be to provide a justification of Christ's words of forgiveness from the Cross?[27] His views here are part and parcel of his general position that an ultimate justification of moral principles is impossible because there are different conceptions of human good and harm, and that it is absurd to look for an ultimate justification of these in terms of some further principle or standard.[28] A full consideration of this position would take us away from the question of forms of life into the realm of ethics. But there is a special case which we must consider, and that is the case of forms of life and language-games which are specifically religious, especially praying, blessing, cursing, and worshipping.

The problem is this: one may agree that it is senseless to ask for a justification of hoping, giving orders, and so forth, and that Wittgenstein is quite right in saying that people think, chat, ask questions, and bring up children because that is what they *do*, and there is no more to be said. But it does seem a sensible question to ask why people pray or worship, and why these language-games are not a universal form of human behavior. Cursing, too, is a language-game which isn't quite as common as it once was, and perhaps Wittgenstein was thinking of examples like this when he said that new language-games come into existence while others become obsolete and get forgotten (P.I. 23). What would happen if religious language-games were to pass away, as witchcraft almost has, at least in the western world? The question is even more pressing when we turn our attention to customs, institutions, and activities (i.e., Wittgenstein's other categories which are relevant here). These certainly do change rapidly and differ from country to country and century to century, and it seems highly relevant to ask in every case why this is so. Of course, Wittgenstein would not deny this; he mentioned customs and institutions to make a different point, namely, that people's concepts, e.g., "price" or "value," are linked to certain practical institutions, like buying and selling. He would say that an investigation of why people buy and sell was not a matter of philosophy so much as of anthropology. But surely it is a highly interesting question for philosophy to ask why certain forms of behavior, linguistic or otherwise, are common to all mankind, e.g., hoping, making up stories or greeting friends, while others are not so, e.g., praying, playing chess, and acting in plays. Changes in such things often have important consequences for our conceptual scheme: for instance, if the institution of punishment were abandoned, as Lady Wootton has advocated, this would affect many of our everyday concepts like "praise," "blame," "reward," "try harder," "responsible," and "atone for."

---

[26] "Wisdom's Gods," *The Philosophical Quarterly*, vol. 19 (1969), p. 29.

[27] *The Philosophical Quarterly*, vol. 19 (1969), pp. 29–31.

[28] D. Z. Phillips and H. O. Mounce, *Moral Practices* (London, 1970), pp. 58–59, 106–112.

V. Two Strands in Wittgenstein's Thought

I think that Wittgenstein's thought here is much more complex than his followers realize and that it moves in two different directions. He seems to envisage two possible answers to the question of why we have the language-games and forms of life which we do: sometimes he just says "this is what we do," sometimes he says that we must connect what we do with facts about the world and about human nature. Of course both these replies are appropriate in different circumstances, but Wittgenstein does not seem to be entirely clear about when each is appropriate. Let us try to distinguish these two strands in his thought:

(1) Sometimes he seems to favor what might be called Relativism or Conventionalism. He imagines that a tribe might be brought up never to express their feelings, so that they regarded complaining as bad as we regard shamming, and goes on to suggest that people with a different education from ours might have completely different concepts (*Zettel*, 383, 387). He imagines a people who give what money they please for goods, and get back what the merchant pleases; this seems entirely pointless to us, but then religious rituals and the coronation of a king are pointless, too (R.F.M. I. 152). He thinks there is little that one could say to people who think they go to the moon and back in their dreams by a special route, who think that kings can make rain, who refuse to rely on calculations or who prefer to consult oracles instead of using physics (*On Certainty*, 106, 132, 212–217, 609). Reasoning only goes so far, after which we use persuasion, like missionaries among natives (*ibid.* 612). What men consider reasonable or unreasonable changes from one age to another; some intelligent people believe in the story of the creation in the Bible, while others don't—yet they both know about the grounds for this (*ibid.* 336). A language-game is unpredictable, since it isn't based on grounds—"It is there—like our life" (*ibid.* 559) (and yet in the previous section Wittgenstein takes a more empirical line when he says that although it is possible that water may not behave in future as it does now, still it always has behaved thus up till now, and this has become fused into the foundations of our language-game).

It is just this strand of Wittgenstein's thought which has been used by thinkers like Peter Winch and D. Z. Phillips. The former, for instance, argues that different modes of life have different "criteria of logic" (whatever that means) and forms of intelligibility.[29] He applies this to magic in his essay "Understanding a Primitive Society," arguing that terms like "reality" and "rational" are context dependent, so that even witchcraft or magic may constitute a "coherent universe of discourse like science."[30]

It is difficult to know how far Wittgenstein himself would sympathize with the use made of his work by Winch and others. It does seem consistent with some of his remarks which I have quoted. On the other hand, as may have become apparent through my linking of Wittgenstein's remarks about forms of life to his thought about surroundings, natural history, etc., there is another strand in his thought which must be taken into account, one which is much more empiricist and pragmatist in its direction.

---

[29] *The Idea of a Social Science* (London, 1958), pp. 100, 102.

[30] *American Philosophical Quarterly*, vol. 1 (1964), p. 309.

(2) Very often Wittgenstein seems to suggest that language-games and forms of life aren't just "given," but that they are closely connected with both facts about the world and about human behavior. Even when things are a matter of convention, the convention may be determined by our situation, and requirements—this is what I take to be the meaning of his puzzling remark that although essences are a matter of convention, there may be a *deep* need for a convention (R.F.M. I. 74).

The facts about human nature which I have in mind here are "ways in which we make comparisons and in which we act" (R.F.M. V. 18), ungrounded behavior (*On Certainty*, 110),[31] and parts of our natural history like walking, eating, and bringing up our children (P.I. 25,467). I have already quoted many passages in which Wittgenstein shows the intimate connection between language and the activities and interests of our life. As regards the "surroundings" or external situations which are to be related to concepts, forms of life, and language-games, he gives many examples too: legal concepts presuppose "a scaffolding of facts," particularly a normal human background, for it would be impossible for a judge to decide whether a being deviant from ordinary human beings had done something with evil intent (*Zettel*, 350). Mathematical procedures are sui generis, but "the facts of daily experience" give them their importance (R.F.M.V. 3). Measuring is a technique, but the technique is only made possible and useful by certain physical and psychological facts (R.F.M. V. 1): if confusion supervened, e.g., if rulers tended to alter in length, measuring and calculating would lose their character and their point (R.F.M. II. 74–75; V. 1–2).[32] Expecting, loving, and hoping arise only in certain surroundings and situations (P.I. 581, 583). Induction is made *logically* possible by the regularity of events (*On Certainty*, 618); this is a particular case of the general point, made in the previous section, that the possibility of a language-game, and of our sureness about it, is conditioned by certain facts. There is also a logical link between "repentance" and "past," for it is not a mere contingent empirical fact that "Only someone who can reflect on the past can repent" (*Zettel*, 519). Presumably there is a similar logical link between the concepts of "hoping" or "expecting" and "the future."

We can also work out the nature of the dependence of our forms of life and language-games on facts about the world by imagining what would happen if the world were different from what it is: "If we imagine the facts otherwise than they are, certain language-games lose some of their importance, while others become important" (*On Certainty*, 63).[33] If houses starting turning into steam or cattle speaking, then would I have any right to say "that's a house"? (*ibid.* 513). If water started behaving differently, our language-games would presumably change (*ibid.* 558). If lumps of cheese tended to shrink sometimes when put on the scale, then fixing the price by the turn of the scale would lose its point (P.I. 142).

The facts which affect our language-games and concepts are often very general ones, so that we tend not to notice them. "What we have to mention in order to explain the significance, I mean the importance, of a concept, are often extremely general facts of nature: such facts as are hardly ever mentioned because of their great generality" (P.I., note on p. 56). Again, if such facts were different, we could imagine people forming concepts different from the usual ones (P.I. p. 230); just as, if people's interests and purposes were other than what they are, their concepts might be different (*Zettel*, 371–379). (Presumably if people did form different concepts from their present ones, these would still be founded on facts about nature and human life.[34])

I do not at the moment see how we can reconcile these two strands in Wittgenstein's thought. I should hazard a guess that he himself was undecided and that he deliberately presented his thoughts tentatively and dialectically (unlike many of his followers!) I am myself prejudiced against the first strand which I have described, the view that forms of life and language-games are simple "given." A philosophy of religion based on this line of thought invariably runs into the objection that many forms of life, language-games, customs, and institutions pass away, so why shouldn't religious ones do so too? Many religions have passed away, while rituals and liturgies become obsolete and get replaced within a religion (it often seems that all liturgies and religious ceremonies appear strange to many people today). Moreover, as I have already suggested, we need to explain the fact that religious forms of life are, at least at the present time, a minority interest, whereas most of the forms of life and language-games mentioned by Wittgenstein himself, e.g., "hoping," "reporting events," or "giving orders" are universal human behavior. Furthermore, if religion is *merely* a given form of life, how do the experiences of conversion and loss of faith fit into the picture?

---

[31] Cp. *On Certainty*, §204: "it is our *acting*, which lies at the bottom of the language-game."

[32] Though one *might* use elastic rulers in certain circumstances, e.g., if one wants to treat customers differently. (R.F.M. I. 5, V. 12).

[33] A passage which has great significance for religious belief: if it were found that Jesus Christ never existed or that his character was significantly different from that portrayed in the Gospels, most Christian prayers and practices would lose their point.

[34] Wittgenstein seems to speak with a divided voice on the question of concept formation. He says that a different education from ours might produce different concepts (*Zettel*, §387). But he also speaks of concepts forcing themselves on us (P.I. p. 204), and of their dependence on certain very general facts of nature (P.I. p. 230).

I think, therefore, that we should try to follow the more "Empiricist" and "Pragmatist" line in Wittgenstein's thought, seeking to show how religious "forms of life" are to be connected with facts about the world, mankind and its history, and with human culture, interests, and purposes. This is not to say that I want to explain (or explain away) these forms of life, for Wittgenstein says "The question is not one of explaining a language-game by means of experience, but of noting a language-game" (P.I. 655).[35] Rather the task is one of describing them, and a full description will involve reference to its context. We cannot *explain* the phenomenon of hope, since it is a basic fact of human life; but we can connect it with other facts, e.g., that people want certain things, that they cannot predict the future, and so on. Similarly with repentance or forgiveness: these are "proto-phenomena" which cannot be explained, but they can be connected with other phenomena like our feelings of guilt or desire to make amends. In other words, they are responses to experience.

VI. Conclusion: Religion as a Response to Experience

Following this line of thought, we would say that religion is a response to experience, or, better, a collection of responses. Christianity, for instance, is a set of forms of life, customs, activities, etc., characterized by the response of love of God and our fellow men in and through Christ.

Clearly this approach, though on the right lines, still leaves us with many serious problems: if loving Christ involves following his teachings, then we become faced with difficult and complex theological issues. And if we join a church, we shall find that our desire to love and follow Christ involves responses and practices of many different kinds. Even a comparatively simple response like "hope" becomes very complex when we are speaking of the theological virtue of Hope, for we have to think of God's promises in the Old Testament, of Christ's words about the future life and the Kingdom of God, of the Church's teaching about Christ's Resurrection, the possibility of grace, and so on. Similarly with "love of God" or "faith in Christ": it is not that they do not deserve the description "love" or "faith," but that we have to draw on a very wide background if we are to explain the concepts fully. They are indeed responses, but ones which depend on some very elusive aspects of life, and ones which have had a long history.

These comments bring us back to our original difficulty in classifying religion as a "form of life": it is too simple a model to account for a complex and historically evolving system of dogma, ethics, and ritual like Christianity. In particular, it is unclear how doctrines relate to forms of life, and this difficulty is, of course, especially grave in the case of metaphysical doctrines about a transcendent being. We may say that worshipping God is a "form of life," but it has no point unless what is worshipped actually exists and is indeed worthy of worship.[36] Now many Christian doctrines are attempts to explain God's nature, activities, and relation to the world; and such doctrines have helped to form the responses of ordinary religious believers, as well as defining the conditions of membership of the Church.[37] A change in doctrine often produces changes in religious practice: for instance, the 16th century Reformers abandoned the practices of praying to the Saints and for the dead. Moreover, those who formulated the doctrines were attempting to convey facts—difficult though it be nowadays to recapture their sense, let alone to decide on their truth.

---

[35] Again, compare *On Certainty*, *op. cit.*, §617, "Indeed, doesn't it seem obvious that the possibility of a language-game is conditioned by certain facts?"

[36] R. F. Holland brushes aside this difficulty by saying that it is not clear what it means to say that God is a being, and he asks what difference does a "special existent" make to the practices of prayer and worship. One may admit his first point and yet still reply to his question "A lot!" How will Holland distinguish superstition from true religion? (See his "Religious Discourse and Theological Discourse," *Australasian Journal of Philosophy*, vol. 34 (1956), especially p. 162.)

[37] See the reply to Holland by J. M. Cameron in *Australian Journal of Philosophy*, vol. 34 (1956), especially pp. 206–207.

The general moral seems to be, therefore, that we need to combine the kind of approaches to philosophy of religion which I have been considering in this paper with a more traditional one. Thus an account of the "form of life" of worship requires a doctrine of God and, in the case of Christianity, a Christology. This is not to say that I regard religious doctrines as mere deductions from a body of facts or that I am looking for a simple pattern of justification. Clearly we must distinguish, for example, the historical facts of Christ's life and our theological interpretation of them. But in the case of Christianity I do think that there is some causal and logical connection between the present responses, activities, and so on of believers, and certain historical events; any complete description and understanding of Christian "forms of life" must take such connections into account. Moreover, any justification of Christian beliefs will have to start from this point.

Darwin College, Cambridge

Received March 12, 1971

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 2, April 1972

V. THE ANALOGICAL INFERENCE TO OTHER MINDS

A. HYSLOP and F. C. JACKSON

THE argument by analogy for other minds (AA) has been widely supposed to involve *conceptual* confusion. It has also been widely supposed to be, simply, a bad inductive argument. In this paper we shall be concerned to defend AA from the latter charge.

AA has been held to be a bad inductive argument because, (i) it proceeds from *one* case, and (ii) it has, of *logical necessity*, an uncheckable conclusion. (Given (i) and (ii) it follows, of course, that it is an argument which, not only in fact but necessarily, is from one case.) We shall call objections to AA which turn on the fact that AA has an uncheckable conclusion, A-type. We shall call objections which turn on the fact, whether necessary or not, that AA proceeds from one case, B-type. Of course, not all objections to AA *qua* inductive argument can be unequivocally classified as A-type or B-type. Such objections that we consider below have been classified according to their general character.

I

In this section we criticize a representative selection of A-type objections. But before doing this we argue that a reply to A-type objections, sufficiently widely discussed to be called "the standard reply," misses the point at issue.

The standard reply to the objection that AA is an analogical argument from the observable to the logically unobservable is that there are perfectly acceptable analogical arguments having just this characteristic; namely, analogical arguments from the present to the past.

The crucial error in this reply is that an advocate of AA cannot allow that the past is unobservable in the sense in which he holds that other minds are unobservable, because the sample class from which AA constructs the inductive generalization to other minds *itself contains past events*; namely, *remembered* correlations between behavior and mental states. Therefore, when it is said, as it so often is, that AA proceeds from what we are (directly) aware of, this must be understood as "what we are, or remember having been, aware of." But it is true of any past event that it is logically possible that we should remember being aware of it, thus a defender of AA must allow that we may be aware of the past in the relevant sense of "aware."

CRITICISM OF A-TYPE OBJECTIONS

The various A-type objections that we shall consider have in common the claim that

(P) No analogical (inductive) argument which proceeds from observed facts to conclusions about what cannot, logically, be observed is a good analogical argument.

The differences between them lie in the reasons, if any, offered in support of the principle (P) embodied in this claim.

Two points of clarification need to be made about (P). The first is that, as the mental states of others are, strictly speaking, not unobservable (they are observable by others), (P) should be read as prohibiting a particular person's arguing analogically from what *he* as propounder can observe to what *he* as propounder cannot observe. We shall follow the usual practice of keeping this implicit, rather than making it explicit, in our discussion.

The second point of clarification concerns the relevant sense in which AA is an argument to what cannot be observed. It is commonly asserted that what generates the problem of other minds is the impossibility of being *directly aware* of the mental states of others. This is a mistake. First, it is not at all clear that it is impossible to be directly aware of another's mental state. Secondly, and more importantly, being directly aware of someone else's pain would not, in itself, tell me that that person was in pain, because I should still require grounds for supposing that what I was aware of *was* someone else's pain (as well as my own). These grounds would have to lie in that person's behavior (or brain states, etc.). Therefore, the relevant sense in which AA is an argument to what cannot be observed is that it is an argument to what cannot be *directly known*. However, nothing in what follows turns on the precise sense in which others' mental states cannot be directly checked, verified, observed, or whatever.

Objection A1

The first A-type objection that we consider consists, in effect, of the bald assertion that AA violates (P). Thus, it is urged that AA is a bad argument because it purports to establish, on analogical grounds, conclusions about items, namely the mental states of others, which cannot be directly observed.

Clearly A1 is a good objection only if one is entitled to assert (P) without offering specific reasons for (P); that is, it is a good objection only if (P) is, in some sense, self-evident. We shall argue, against A1, (i) that not only is (P) not self-evident, but that, in the absence of positive reasons, (P) should be taken to be false, and (ii) that a possible reason for supposing (P) true involves confusing the possible with the actual.

(i) It is clear for two reasons that (P) is not self-evident. First, the fact that an inductive argument from premisses *X* to conclusion *Y* violates (P) shows, *in itself*, only that *Y* is such that evidence for it additional to *X* of a direct kind cannot be obtained. Now to say that *Y* is such that certain kinds of additional evidence for it cannot be obtained is to say something which, on the face of it, is quite different from saying that *X* does not support *Y*. Therefore, no claim that the former implies the latter can be regarded as self-evident.

Secondly, there are perfectly acceptable inductive arguments which, while not violating (P), violate a principle sufficiently close to (P) for the onus of proof to lie with a defender of (P). Here is an example of such an argument.

Helium, like all the elements, has a characteristic spectrum. The nature of this spectrum was determined by experiments carried out, of course, *on earth*. It was then observed (actually, it was remembered) that light from the sun displayed this spectrum (along with hundreds of others), and from this it was inferred that helium was present on the sun. Now the argument just outlined is an argument which proceeds from the fact that certain observed samples of helium on earth have a certain spectrum, to the conclusion that the helium on the sun, which has not been observed, also has this spectrum. Moreover, it is an argument from the observed to the *unobservable*, for it is *impossible* to observe the helium on the sun. Now, of course, this impossibility is empirical, not logical, so there is a difference between the argument for helium on the sun and AA. At the same time, it seems clear that this difference is not *obviously* relevant, therefore the onus of proof lies with an advocate of (P); that is, (P) is not self-evident. Indeed, the helium argument shows not just that (P) is not self-evident, but that (P) should be taken as false in the absence of positive reasons for (P).

(ii) Our argument for the claim that a possible reason for supposing (P) true involves a confusion between the possible and the actual follows Carl Wellman, "Our Criteria for Third Person Psychological Sentences," where he says:

> . . . it is easy to see why one is reluctant to accept indirect (analogical) justification where direct justification is absent. There is always a risk in indirect justification, for it is essentially the application of a generalisation based upon past experience to a new situation . . . Although the possibility of error cannot be denied, it is reassuring to know that there are tests that could determine whether or not an error had been made. And if there were no possibility of direct verification, then one could never be quite sure.
>
> All this . . . is true . . . [it] is also irrelevant. Until one has actually obtained this direct verification, it can hardly be included in one's reasons for holding the statement to be true. . . . Whether a person might have had or could obtain another (direct) kind of justification is irrelevant to the validity of the justification he actually possesses.[1]

Wellman takes the argument just quoted to refute objections to AA like A1. It does not. What it refutes, and this is the point we wish to make, is any argument for A1 along the lines that the problem with any indirect analogical argument (namely that it involves a generalization from what has been directly verified to what has not been directly verified) is resolved by appealing to the *possibility* of directly verifying or failing to verify what one is indirectly arguing to.

In addition it is worth noting that those who find reassurance in the possibility of directly *verifying* the conclusion of some analogical argument appear to have overlooked the fact that the possibility of directly verifying the conclusion of AA goes hand in hand with the possibility of directly *refuting* the conclusion of AA. (Though it is not true *in general* that direct refutation goes hand in hand with direct verification: witness "I am immortal.") Because AA violates (P), AA is an analogical argument such that direct observational evidence in *favor* of its conclusion is impossible. But, it is equally the case that because AA violates (P) it is impossible to have direct observational evidence *against* its conclusion. Therefore, if the impossibility of direct evidence in favor of AA's conclusion counts against the soundness of AA, then, by parity of reasoning, the impossibility of direct evidence against AA's conclusion counts in favor of the soundness of AA. Therefore, the two balance out and leave the soundness of AA untouched.

---

[1] C. Wellman, "Our Criteria for Third Person Psychological Sentences," *Journal of Philosophy*, vol. 58 (1961), pp. 281–293 (see pp. 292–293).

We now turn to various A-type objections which differ from A1 in that specific reasons are offered in support of the claim that, because AA violates (P), it is unsound.

Objection A2

This is put by Don Locke thus

> . . . if, in an ordinary case, we are doubtful about whether an analogy holds we can check up and see whether it does, but the whole point about other minds is that we can never test whether the analogy does hold.[2]

In effect, Locke's suggestion is that the reason why the violation of (P) by AA matters is that, if we do have doubts, it is impossible to resolve them.

Locke does not make it clear whether the doubt he is discussing is about the truth of the conclusion of AA, or about the inductive soundness of AA. If we understand being doubtful about whether the analogy holds as being doubtful about whether the cases are sufficiently analogous for an analogical argument to be justified then A2 rests on a false premiss. The analogy in question is that between one's own behavior and the behavior of others, and it is just false that one cannot resolve doubts by checking up on just how analogous one's own behavior is to that of others. On the other hand, if we understand being doubtful about whether the analogy holds as being doubtful about whether others are analogous to ourselves in having minds, then A2 is either irrelevant or a *petitio principii.* It is irrelevant if the doubt in question is *psychological.* The fact that if one were to have a certain kind of doubt this doubt could not be removed has nothing to do with whether one *ought* to have this doubt. On the other hand, if the doubt in question is *rational.* A2 is a *petitio principii.* It is the contention of defenders of AA that it is *not* rational to doubt whether others have minds just because AA is a good argument. Therefore, it is blatantly question begging to assume that it is rational to doubt whether others have minds in advancing this objection to AA.

Objection A3

> The most familiar A-type objection to AA turns on the claim that any acceptable analogical argument must rest on an *independently* established correlation, and that, because AA violates (P), there is no appropriate independently established correlation for AA to rest on.[3]

Despite the widespread acceptance of this kind of objection, it is hard to see any force in it. It is just false that "there could be no way of establishing the necessary correlation between the overt moves and their hidden causal counterparts," because one can establish a correlation in one's own case. It may be replied that a correlation in one's own case between behavior and mental states could only entitle one to argue analogically about one's own mind, and that to argue analogically to the minds of others it is necessary to establish a correlation between the behavior of others and the mental states of others. This reply is, however, quite unacceptable; because it amounts to the claim that if *x* is inferred from *y* then *x* must on at least one occasion have been directly observed to be present when *y* was present. We could not, that is, infer to a party in a shuttered house unless we had on at least one occasion observed directly a party in *that* house, even though we had observed a score of parties in a score of houses other than the house in question.

At this point a modification of A3 might be suggested, *viz.*, to construe A3 so that it turns on the fact that while in normal cases conclusions arrived at analogically can be independently checked, in the case of the analogical inference to other minds no conclusion reached by the use of this method of inference can ever be independently (directly) checked. In short, the modified objection is that AA is a method of inference whose use has never, and can never, be shown to be successful.

Our reply to the modification follows Hampshire in "The Analogy of Feeling" where he says

[2] D. Locke, *Myself and Others* (Oxford, 1968), pp. 49–50.

[3] The *locus classicus* is G. Ryle, *The Concept of Mind* (London, 1949), see p. 52.

> . . . each one of us is constantly able to compare the results of this type of inference [AA] with what he knows to be true directly and non-inferentially; each one of us is in a position to make this testing comparison, whenever he is the designated subject of a statement about feelings and sensations [made by another on analogical grounds].[4]

Malcolm, of course, and his view is widely accepted, thinks Hampshire guilty of an *ignoratio elenchi* (presumably he meant to say *petitio principii*), arguing that

"... the reasoning that he [Hampshire] describes involves the assumption that other human figures *do* have thoughts and sensations; for they are assumed to *make inferences* about me from *observations* of my behaviour,"[5] and, obviously, a defender of AA is not entitled to make this assumption. Now it is not clear whether Hampshire's argument, as he formulates it, makes this assumption; but it *is* clear that it need not make this assumption, because it does not depend on the assumption that other human figures are minded. The argument is hypothetical in character; it is that *if* other human figures make inferences about my mind on the basis of observing my behavior, *then* those inferences, involving analogical reasoning, can be directly tested.

A more extreme modification of A3 is sometimes put in terms of the impossibility of knowing that the method of inference involved in AA is reliable; that is, is a method which regularly yields true conclusions. Thus V. Chappell's imaginary sceptic demands that the method of inference be known to be perfectly reliable, where ". . . to know that a method of inference from x to y is perfectly reliable is to know that x occurs whenever y occurs."[6]

This version of A3 fails for two reasons. First, the objection depends on an illegitimate identification of a sound inductive method of inference with a perfectly reliable method of inference. The whole point about good inductive arguments, as opposed to valid deductive arguments, is that their premisses may be true when their conclusions are false. Secondly, and more importantly, since any defender of AA does, of course, believe that his method of inference is very probably, though not necessarily, perfectly reliable; this version of A3 is a *petitio principii*. Just because a defender of AA thinks AA is a good inductive argument, he holds that we do know that the method of inference in question is perfectly reliable. The objector can only hold that we do not know with high probability that the method of inference is perfectly reliable, *if* he assumes that AA is not a good argument, which is to beg the question at issue.

Objection A4

This is that because AA violates (P), the *observational* evidence, as far as an advocate of AA is concerned, is simply that others do not have minds.

As Alexander puts it:

> In spite of the fact that similar behaviour to that I exhibit, accompanied by my experiences, is exhibited by everyone else unaccompanied, as far as I can observe, by experiences—in spite of the fact that is, that I observe more samples of purposive behaviour *without* experiences than I do samples of such behaviour with experiences—I am yet content to conclude that all those samples are really accompanied by experiences, when I should, on the evidence, it seems, conclude that I am unusual.[7]

Now it is, of course, true that when we fail to observe that something is the case this can generally be taken as counter evidence to the claim that it is the case. Nevertheless, as Plantinga points out:

> There is a difference, in general, between failing to observe the presence of *A*'s and observing the absence of *A*'s (observing that no A's are present).[8]

And where it is impossible to observe the absence of *A*'s then the consequent failure to observe the presence of *A*'s is in no way counter-evidence to the presence of *A*'s. Therefore, since it is not possible to observe the absence of mental states in the case of another human figure, our failure to observe their presence is not counter-evidence to there being such mental states.

Objection A5

This objection is that put forward by A. Plantinga.[9] It is not strictly an A-type objection because it does not turn on (P). It does however turn on a principle akin to (P), namely (A), below.

---

[4] S. Hampshire, "The Analogy of Feeling," *Mind*, vol. 61 (1952), pp. 1–12 (see p. 4).

[5] N. Malcolm, "Knowledge of Other Minds," in V. Chappell, (ed.), *The Philosophy of Mind* (Englewood Cliffs, 1962), pp. 151–159 (see p. 153).

[6] V. Chappell, (ed.), *The Philosophy of Mind* (Englewood Cliffs, 1962), p. 3.

[7] P. Alexander, "Other People's Experiences," *Proceedings of the Aristotelian Society*, vol. 51 (1950/51), pp. 24–46 (see pp. 43–44). Alexander is considering, without definitely endorsing, this objection.

[8] A. Plantinga, *God and Other Minds* (Ithaca, 1967), p. 248.

Plantinga's objection may be set out as follows. "A simple inductive argument for *S*" is defined as an argument of the following form:

Every *A* such that *S* has determined by observation whether or not *A* is *B*, is such that *S* has determined by observation that *A* is *B*. Therefore probably every *A* is *B*.

Plantinga then suggests that

(A) A simple inductive argument for *S* is acceptable only if it is logically possible that its sample class contain a counter instance to its conclusion.

Plantinga's objection is that AA requires certain simple inductive arguments that violate (A). For example, AA requires the inference from

(1) Every case of pain behavior such that I have determined by observation whether or not it was accompanied by pain in the body displaying it, was accompanied by pain in that body

to

(2) Probably every case of pain behavior is accompanied by pain in the body displaying it.

And the inference from (1) to (2) violates (A) because it is logically impossible to get in the sample class a counter-instance to the conclusion. This is logically impossible because I cannot determine by observation that a body does *not* contain a pain; I can only determine that it does not contain a pain of *mine*, which leaves open the possibility that it contains a pain belonging to someone else, as any body, including my own, may contain someone else's pain.

We agree with Plantinga in accepting (A), and agree that the inference from (1) to (2) violates (A). Our objection to his objection to AA is that AA does *not* involve an inference from (1) to (2). Now, it is easy enough to show that AA does not involve an inference from (1) to (2), by, for example, pointing out that any analogical arguer must regard (2) as false in virtue of the phenomenon of pretending. However, it is clear that this does not get to the heart of the matter. What is needed to show Plantinga's objection fails is an inference which, while being of the same general character as that from (1) to (2), does not violate (A). We suggest that the inference from

(3) Every case of pain behavior such that I have determined by observation whether or not it was accompanied by a pain in the body displaying it *which belonged to the person whose body displayed the pain behavior*, was so accompanied.

to

(4) Probably every case of pain behavior is accompanied by a pain which is in the body displaying it and which belongs to the person whose body is displaying it,

is such an inference. It is an inference which constitutes a simple inductive argument, and it does not violate (A), because it is possible to determine by observation that a body does not contain a pain of mine.

A possible reply for Plantinga at this stage, as John Fox suggested to us, is to point out that the step from (3) to (4) does not violate (A), only if we are allowed to talk of *the* person whose body displays the pain behavior in (3). If we allowed (as Plantinga no doubt holds to be possible) that human bodies may have more than one person "attached" to them, and, so, talked merely of *a* person in (3), then a counter-instance to the conclusion would not be possible. A counter-instance would not be possible, and, so, (A) would be violated, because if I determined by observation that my body was displaying pain behavior and that my body did not contain a pain of mine, it would still be possible that my body contained a pain of a person (other than myself) who shared my body with me.

Our counter to this reply is that, though we agree both that the assumption that there is no more than one person attached to my body is not necessarily true, and that the assumption is needed to allow us to replace (1) by (3), and so to avoid violating (A), nevertheless the assumption is *true*. And if the assumption is true, and known to be true, we are entitled to make it in (3). It would of course be circular then to use the analogical argument to support the assumption, but it is obvious in the first place that one could not expect the analogical argument to establish that there is just one person to my body because the analogical argument proceeds from what I am directly aware of, and I am not directly aware of this fact.[10]

---

[9] Plantinga, *loc. cit.*, pp. 245–268. See also his "Induction and Other Minds," *Review of Metaphysics*, vol. 19 (1965–1966), pp. 441–461 and Michael Slote's reply in *Review of Metaphysics*, vol. 20 (1966–1967), pp. 341–360 and Plantinga's reply to Slote in the same issue. We discuss Plantinga's criticism of AA because, apart from being an unusually penetrating one, it seems to us that Slote's reply to it fails.

Objection A6

The last A-type objection we consider is that put forward by Sydney Shoemaker. He says:

> It has often been suggested, of course, that one can discover empirical correlations between *one's own* psychological states and facts about one's own body and behaviour, and that one's knowledge of these correlations can provide a basis for inductive inferences from the bodily and behavioural states of others to assertions about their psychological states. . . . But even if one could discover this, one could have no grounds for supposing that the correlations that hold in one's own case hold in the case of other persons as well.[11]

Shoemaker's reason for holding that "one could have no grounds for supposing that the correlations that hold in one's own case hold in the case of other persons as well" is that anyone who employs AA presumably holds that behavioral and psychological states are only contingently related; and, therefore, that it is contingent that the relation which holds in one's own case holds in the case of others. But, if this is contingent, the analogical arguer can have no ground for supposing it to be the case, because the only empirical evidence he has concerns the relation in his own case. He has not, for example, independently established a correlation between the relationship between physical and psychological in his own case and the relationship in another's case, to sustain the generalization that the relationship is the same in all cases.

Shoemaker's objection raises issues in inductive logic which are beyond the scope of this paper. We shall content ourselves with pointing out that *if* A6 is successful, then *no analogical arguments whatever are successful*; because what Shoemaker says about AA can be said about *any* analogical argument concerning the unobserved. Whenever one argues inductively from all observed *A*s having been *B*s to the unobserved *A*s being *B*s, it will be true both that it is contingent that the unobserved *A*s are like the observed *A*s (otherwise the argument would be deductive, not inductive), and that there is no independently established correlation between observed *A*s and unobserved *A*s to sustain the generalization that observed *A*s and unobserved *A*s are alike (otherwise the unobserved *A*s would be observed, not unobserved).

II

We now turn to B-type objections to AA. Before criticizing them we shall argue, as we did concerning A-type objections, that a reply to B-type objections sufficiently widely discussed to be called "the standard reply," fails.

The standard reply is a reply to the objection to AA that it is a wild generalization from one case, and is, in effect, that the objection commits the *fallacy of equivocation.*

If "one case" means just one occasion of a mental state being correlated with a physical state, then the objection is simply based on a false premiss. It is not true that AA is based on one case in this sense. It is based on the multitude of correlations between mental and physical that one has observed in one's own case. On the other hand, so the standard reply goes, if "one case" just refers to the fact that the multitude of correlations between mental and physical are all in one's own case, then the objection is merely drawing attention to the fact that the members of the sample on which AA is based have a common property; namely, of being correlations in one's own case. And it is quite generally the case that analogical arguments proceed from samples having common properties. For example, in the helium argument discussed before, all the samples of helium were on earth.

It seems to us that this standard reply is quite mistaken. Despite the fact that there is a multitude of correlations in one's own case between mental and physical goings on, there is an important sense in which AA is an analogical argument from only one case, rather than an analogical argument from a number of cases having a common property. Two rather different considerations show that AA must be taken as an argument from one case.

The first turns on the point that it is just false that *every*, rather than most or many, instances of, for example, pain behavior on one's own part are accompanied by a pain of one's own. This means that if we regarded AA as proceeding from the many instances of mental-physical correlation in one's own case, the only conclusion that could be obtained would be: Many instances of behavioral states are associated with mental states. Now this conclusion is just not the conclusion we want AA to yield, because it is consistent with certain human bodies which behave just as one's own does, not having *any* associated mental states. The conclusion we want AA to yield is that for *every* body which behaves much as one's own does, it is true that many instances of its behavioral states are associated with mental states.

---

[10] See H. Castañeda, "Consciousness and Behaviour," in *Intentionality, Minds and Perception*, ed., Castañeda (Wayne State, 1966), pp. 132–133.

[11] S. Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, 1963), p. 168.

The second consideration arises from the fact that we believe that for each living human body there is at least one person such that the behavioral states of that body are accompanied by the mental states of that person. (We also believe, of course, that there is *exactly* one person to each living human body; but in view of our reply to A5, this is not something we should expect AA to establish.) The only analogical argument that could show this would be one generalizing from the fact that for one's own body there is at least one person (oneself) such that the behavioral states of that body are accompanied by the mental states of that person. Now this is, and must be, an inductive generalization from *one* case; therefore, if the analogical argument for other minds is to establish that for each living body there is at least one person whose mental states correspond to the behavior of that body, the argument must be regarded as an analogical argument from one case.

Criticism of B-type Objections

If our comments on the standard reply are correct, then AA is an analogical argument from one case. This is the point on which the major objection we consider below turns; however, we first consider an objection sometimes confused with this objection.

Objection B1

This is that the ratio of examined to unexamined instances is not high enough. Don Locke complains that "[AA] proceeds from the smallest possible basis to the largest possible conclusion; it argues from one solitary instance to, perhaps, an infinity of cases."[12]

This objection involves a misunderstanding about the nature of AA, in that it involves treating AA as a species of *statistical* argument. In fact, as the following argument shows, the soundness of AA is independent of the ratio of examined to unexamined cases.

Suppose I use AA in an attempt to establish that just one other person has a mind, then no complaint can be made about the ratio of examined to unexamined cases because this ratio will be very much higher than in the vast majority of cases of inductive generalization. But, clearly, this one other might be *any* other. Therefore, if it is reasonable to use AA to show that one other person has a mind, then, for any other person, it is reasonable to use AA to show that he has a mind; provided, of course, that he is behaviorally sufficiently like myself. The reason why the ratio of examined to unexamined cases is irrelevant to the soundness of an analogical argument is that the soundness of an analogical argument depends on the relevant analogy between the examined and the unexamined cases. If the analogy between the examined cases and some given unexamined case is sufficient to sustain the inference, then, by parity of reasoning, in *any* unexamined case where the analogy holds, regardless of how many of these there are, we are entitled to make the inference in question.

Objection B2

The last B-type objection that we consider is that put by Norman Malcolm. He says

". . . this [AA] would be very *weak* inductive reasoning, based as it is on the observation of a single instance."[13] This objection is distinct from B1 because it turns, not on the ratio of examined cases to unexamined cases being low, but on the fact that there is only one examined case. As such, it is *prima facie* a good objection because it is indisputable that we often require arguments from analogy to be based on more than one instance. (However, this objection should not be confused with the objection: "How do I know that it is not my individual peculiarity to feel angry when I frown?"[14] One may enquire of any inductive base, how do I know that it is not a peculiarity of the sampled cases to be as they are? Since it can be raised against *any* analogical argument it is not an objection to AA as such.)

Although we often demand that an analogical argument be based on more than one instance, nevertheless there are analogical arguments which we take to establish their conclusions with very high probability which proceed from only one examined case. Our reply to B2 will be to argue that AA is of the latter kind. Our argument will proceed via a discussion of an analogical argument which, as it stands, requires to be based on more than one instance.

---

[12] Locke, *loc. cit.*, p. 49.

[13] Malcolm, *loc. cit.*, p. 152.

[14] J. Day, *Inductive Probability* (London, 1961), p. 64.

Suppose I drop an egg on the floor and it breaks. Am I entitled to infer that the next egg dropped will break? Obviously, we should all be happier with such an inference if it were based on a number of cases of eggs that had been dropped and had broken. Why should we be happier? Because we should have a statistically more significant sample? Surely not, because the total number of eggs there are is so large that a few more cases will not yield any significant increase in the percentage of eggs tested. Surely the answer is that if we carry out more tests we rule out the possibility of the breaking in the case of the tested eggs being due to something other than their being dropped. What appears to follow from this is that if we know, somehow or other, that the first egg broke just because it was dropped, then the additional tests would not have been necessary. We shall put this point more generally. Suppose that we are arguing from the fact that $x_1, \ldots x_n$ are both *F* and *G*, and that *y* is *F*, to the conclusion that *y* is *G*. When do we need *n* to be large? The suggestion we are making is that we need *n* to be large just when (i) a large sample will provide us with information about the (causal) relation between an *x* being *F* and being *G*, (ii) this information is necessary for the inference to *y* being *G*, and (iii) we do not already have this information. The important point is (iii). If we already have the required information the sample need not be large.

We can now apply this suggestion to show that AA does not need to proceed from more than one case. In the one case from which AA proceeds we know, without appeal to a larger sample, that the possession of a mind is the operative cause of the displaying of behavior of the kind which in others is then the basis for supposing that they have minds. I know in my own case that my body behaves as it does because I have a mind; and I know this independently of what I know about the minds of others, because I would know this even if I were the only person in existence. However, the information that, in the sampled case (my own), a mind is the operative cause is all that a larger sample could give me relevant to the inference from the behavior of another to the mind of another. Therefore, as I already have the only relevant information a larger sample could yield, a larger sample is unnecessary. To ask where does this information come from would be beside the point. We have it, and we have it, as noted, independently of our knowledge of other minds, and so, there is no circularity in using this information to support AA.[15]

In our view what has bedeviled discussion of B2 is the feeling that the important point arising from the fact that AA is based on just one case is that one is left without direct evidence against this one case being unique. But what good, if it were possible, would enlarging the sample on which AA is based do? It would provide direct evidence that one's own case was not unique, of course, but how would this help? One would still have no direct evidence that the sampled cases were not unique. It is a necessary truth that no analogical argument, no matter how large its sample, provides direct evidence that its sample is not unique. Therefore, sample size must have some other role to play in analogical arguments. We have tried to argue that its role is to provide certain causal information, and that this role is not needed in the case of AA because we already have this information.

In arguing this, we are not denying that the possibility of the sample being unique raises a serious problem. "How does one know that the cause in the sample case is the cause in the unsampled cases?" is a perfectly reasonable question. But, as we have in effect already noted, it is a question that can be raised against *any* analogical argument. The fact is that we know that the cause of certain behavior in our own case is mental. To suppose, in the absence of any plausible alternative cause, that the operative cause in the case of others is different would be to indulge in a fantastic piece of *a priori* speculation. Why would this be fantastic? Perhaps Occam's Razor and the undoubted *predictive* power of the other minds hypothesis provide part of the answer to this question; but, at any rate, the answer to this question will be the answer to the problem of induction for it will tell us when it is reasonable to suppose that a sample is like a population, and, fortunately, the problem of induction has not been the problem with which this paper has been concerned.

---

[15] Our emphasis on the causal role of mental states with respect to behavior commits us to holding that epiphenomenalists and parallelists cannot use AA. This is as it should be; if one holds that a person would behave exactly as he does even if he had no mental states, then one can hardly regard his behavior as providing a good inductive ground for assertions about his mental states. In this respect we are following J. S. Mill who makes it clear that the causal role of mental states is central to his version of the analogical argument for other minds: see his *An Examination of Sir William Hamilton's Philosophy*, 3rd ed. (London, 1867), p. 237.

The problem this paper has been concerned with is whether, given that there are some sound analogical arguments, there is any good reasons for denying that AA *qua* analogical argument is such an analogical argument. We conclude that there is not.[16]

La Trobe University

Received February 9, 1971

---

[16] This paper has benefited from discussions with colleagues, particularly R. J. Pargetter and R. J. Pinkerton.

American Philosophical Quarterly
Volume 9, Number 2, April 1972

VI. LOGICALLY NECESSARY AND SUFFICIENT CONDITIONS FOR IDENTITY THROUGH TIME

JACK NELSON

IN recent years a considerable interest has developed in the identity conditions for material objects picked out as existing at different times, e.g., in what is required for the briefcase I now carry to be numerically identical with the briefcase I bought in Edinburgh in 1965. The thesis I wish to defend is that we neither have, nor are likely to find, any interesting logically necessary and/or sufficient conditions for the identity of material objects picked out as existing at different times. I shall take as my chief stalking horse the view that spatio-temporal continuity either itself constitutes or at least provides the basis for such a set of conditions.

I

It is widely held that what are commonly referred to as Leibniz' Law and the Identity of Indiscernibles Principle together give us a logically necessary and sufficient condition for numerical identity. This requirement can be formalized as follows:

$$(1)\quad (\forall x)\ (\forall y)\ (\forall F)\ [(Fx \equiv Fy) \equiv (x = y)].$$

It is not entirely clear whether we should read schema (1) as saying any x and any y are identical if and only if they have all of their properties in common, or if and only if they are indiscernible, or if and only if whatever is true of x is true of y, and vice-versa, . . . or in some other way. Moreover it has been claimed that intentional contexts and the possibility of radically symmetrical universes provides counterexamples to Leibniz' Law and to the Identity of Indiscernibles Principle, respectively. However, it is not my purpose either to defend or to attack Leibniz' Law and the Identity of Indiscernibles Principle. What I do want to argue is that if schema (1) does in fact give us a logically necessary and sufficient condition for the identity of all kinds of things whatsoever, as I think it does, that condition is terribly uninteresting and uninformative, at least as applied to material objects picked out as existing at different times.

N. L. Wilson has noted that when we ask "When are two individuals identical?" "We are told that two individuals are identical if and only if they have all their properties in common. The difficulty with this answer is," Wilson continues, "that it merely states the syntactical properties of the identity sign but gives us no help in deciding in a given case whether to declare two individuals identical."[1] The force of Wilson's comment is not hard to see. Suppose we pick out a red table as existing at time t, call it 'A', and a green table as existing at time t', call it 'B', where $t \neq t'$. A naive reading of schema (1) will here lead us wildly astray. That is, we might reason that since A is red and B is green they do not have all of their properties in common (are not indiscernible, not everything true of A is true of B, and vice-versa, etc.), and hence cannot be numerically identical. This result is clearly objectionable, one and the same table can clearly be different colors, *at different times*. And this fact need not be seen as constituting an exception to schema (1), so long as we are mildly careful about how we specify the properties of (or what is true of) things. The easiest procedure to adopt here is that of "time-flagging" properties. We say, for example, that A has the property, not of being red, *simpliciter*, but rather of being red at t. Likewise, B has the property of being green at t'. And since $t \neq t'$ these are compatible properties. But, and this is the point, there is nothing in schema (1) which tells us how to apply this principle to particular cases, or kinds of cases, nothing which tells us not to read that principle in the naive way detailed above.

In other words, Leibniz' Law and the Identity of Indiscernibles Principle seem to impose syntactical, not substantive, requirements for numerical identity. It is beginning to look as if we can call whatever things we want identical, so long as we find a way of specifying the properties of those things which allows us to ascribe all the properties of each to the other in a manner which is consistent with the facts, as those facts are described by our convention for specifying properties. But this result is premature. Though Leibniz' Law and the Identity of Indiscernibles Principle are uninformative and unrestrictive, at least as applied to material objects picked out as existing at different times, there may be other, more restrictive and therefore more informative, requirements for the identity of particular kinds of things, e.g., for the identity of material objects picked out as existing at different times.

---

[1] N. L. Wilson, "Space, Time, and Individuals," *The Journal of Philosophy*, vol. 52 (1955), p. 592.

It should be clear that having one logically necessary and sufficient condition for the holding of some relation does not preclude there being more specific conditions for that relation's holding. For example, a logically necessary and sufficient condition for any person, *x*, being a relative of any person, *y*, is that *x* be related to *y* by blood or marriage. But there are also a host of specific logically necessary and sufficient conditions for *x*'s being related to *y*, in specific ways. Thus *x* is an uncle of *y* if and only if *y* is the son or daughter of one of *x*'s (or *x*'s spouse's) siblings. Now precisely because schema (1) is so nearly vacuous, at least as applied to material objects picked out as existing at different times, it is plausible to think there may be specific logically necessary and/or sufficient conditions for the identity of particular kinds of things, e.g., numbers, concepts, material objects. That there are, in fact, no such specific conditions for the identity of material objects picked out as existing at different times is thus a substantive claim, a claim which requires supporting argumentation. I shall now try to provide this argumentation.

II

Some years ago B. A. O. Williams put forward the following principle, let us call it "the Williams Principle":

> [I]dentity is a one-one relation, and . . . no principle can be a criterion of identity for things of type *T* if it relies only on what is a logically one-many or many-many relation between things of type *T*.[2]

The reasoning behind the Williams Principle is simply this. Identity is a logically one-one relation. That is, it is a matter of logic, and not just of fact, that if *x* is identical with *y*, then *x* cannot also be identical with *z*, unless *y* is also identical with *z*.[3] This being the case, any requirement which constitutes a logically sufficient condition for identity must itself rest on a logically one-one relation. (If it does not it cannot be a *logically* sufficient condition for identity.)

The Williams Principle, when taken as a principle about logically sufficient conditions, rather than about criteria, in the sense of general but violable principles for use in determining when specific things are identical, is clearly correct. It follows that if there is any requirement, or set of requirements, which constitutes a logically sufficient condition for the identity of material objects picked out as existing at different times that requirement or set of requirements must rest on a logically one-one relation.

It is, I suggest, fairly easy to see that unless we appeal to spatio-temporal continuity it is extremely unlikely that we shall find any relation between material objects picked out as existing at different times which is non-trivially a logically one-one relation. It will not suffice, for example, to say the relation we want is that of having all time-flagged properties in common, for this is just the empty requirement of schema (1) rephrased. What we want to know is when such objects do have all of their time-flagged properties in common. Nor will it help to appeal to purely general or descriptive properties (non-time-flagged, non-spatial, and non-relational properties). For having all such properties in common is clearly not a logically one-one relation; coexistent but non-identical objects can be indiscernible with respect to such properties.

The result of the foregoing is just this: it looks extremely likely that if there is a non-trivial logically sufficient condition for the identity of material objects picked out as existing at different times that condition will have to include the requirement of spatio-temporal continuity. Hence if it can be shown that spatio-temporal continuity is not, and cannot form the basis of a logically one-one relation, or alternatively if spatio-temporal continuity turns out to be a relation which material objects picked out as existing at different times virtually never satisfy, the hope of finding an interesting, non-trivial, logically sufficient condition for the identity of material objects picked out as existing at different times will have been dispelled. And if no logically sufficient condition more interesting than that provided by the Identity of Indiscernibles Principle is forthcoming, there seems to be little reason to think logically necessary conditions more interesting than that provided by Leibniz' Law are forthcoming.

---

[2] B. A. O. Williams, "Bodily Continuity and Personal Identity: A Reply," *Analysis*, vol. 21 (1960), pp. 44–45.

[3] Peter Thomas Geach holds that identity is a relative relation, and hence has to deny that identity is a logically one-one relation. But Geach is just wrong about the relativity of identity. See "The Same F," John Perry, *The Philosophical Review*, vol. 79 (1970), pp. 181–200, and my paper "Relative Identity," *Nous*, vol. 4 (1970), pp. 241–260.

Before going on to investigate, in detail, the relation of spatio-temporal continuity one important possible objection to the above line of argumentation must be met. This objection is that we have overlooked one plausible way of forming an interesting logically sufficient condition for the identity of material objects picked out as existing at different times which does not rely on spatio-temporal continuity. Specifically, it has been suggested that we can form a logically sufficient condition for the identity of such objects by centering on some plausible requirement for identity, such as a similarity in descriptive or purely general properties (or, in the case of persons, similarity in purported memory claims, or personality traits, etc.) and adding the proviso that satisfaction of one, or several, of these requirements is a logically sufficient condition for identity in, and only in, those cases in which the objects satisfying the requirement(s) in question do, as a matter of fact, stand in a one-one relation. This seems to be the line taken by Robert Coburn in his article "Bodily Continuity and Personal Identity"[4] and is explicitly the line taken by Richard Gale in his more recent note "Personal Identity and Bodily Continuity."[5]

One is tempted to complain about the *ad hocness* of such sufficient conditions as the above: *x* is numerically identical with *y*, we are told, *provided* no object other than *x* bears the relation *x* bears to *y*. . . . But more importantly such moves, which easily lend themselves to the language of "normally sufficient" conditions, and "criteria" (in some Wittgenstinean sense) seem, to this author at least, to represent an implicit abandoning of the search for logically sufficient conditions. That is, the core of these positions seems to be just the position I am arguing for, namely, that the only logically sufficient and/or necessary conditions for the identity of material objects we are likely to find are those on the level of generality and uninterestingness of Leibniz' Law and the Identity of Indiscernibles Principle. Beyond this we can only try to ferret out the general, but violable, principles we do as a matter of fact use in deciding what objects are identical with what objects, i.e., in deciding what pairs of objects picked out as existing at different times do, and what do not, satisfy the requirement of schema (1).

Perhaps there is more to talk of criteria and normally sufficient conditions than is allowed above. But at any rate before we give ourselves over to the task of eliciting the general, but violable, principles we use in deciding cases of identity, whatever they are to be called, it is important to see that no stronger requirements for identity can be obtained. And this has not yet been done. For it has not yet been shown that spatio-temporal continuity neither is, nor can form the basis for, a logically one-one relation between material objects picked out as existing at different times.

III

B. A. O. Williams, who clearly thinks spatio-temporal continuity is a logically necessary condition for the identity of material objects picked out as existing at different times, and may think it is a logically sufficient condition,[6] expresses some doubts about spatio-temporal continuity's being a logically one-one relation. Fission is characteristic of amoebae, and both the amoebae which result from an instance of fission do seem to be spatio-temporally continuous with their common "parent" amoeba. And we can, Williams claims, "imagine a man splitting, amoeba-like, into two simulcra of himself." Williams' response is to suggest that "in a case of fission, such as that of an amoeba, the resultant items are not, in the strict sense, spatio-temporally continuous with the original."[7] Gale, on the other hand, takes cases of fission as showing that spatio-temporal continuity is not a logically one-one relation.[8] Both responses seem hasty and *ad hoc* in the absence of any account of the "strict" sense of spatio-temporal continuity. But perhaps we can find such an account.

N. L. Wilson is among the philosophers who hold that spatio-temporal continuity is both a logically necessary and a logically sufficient condition for the identity of material objects picked out as existing at different times. His account of the identity conditions for such objects, and thus of spatio-temporal continuity, is as follows:

---

[4] Robert Coburn, "Bodily Continuity and Personal Identity," *Analysis*, vol. 20 (1960), pp. 117–120.

[5] Richard Gale, "A Note on Personal Identity and Bodily Continuity," *Analysis*, vol. 29 (1969), pp. 193–195.

[6] This no longer seems to be Williams' position, at least not as regards personal identity. See his recent article "The Self and the Future," *The Philosophical Review*, vol. 79 (1970), pp. 161–180.

[7] Williams, "Bodily Continuity and Personal Identity: A Reply," *op. cit.*, pp. 47–48.

[8] Gale, *op. cit.*, p. 194.

> $(\iota x)[P_i - t_i(x)] = (\iota x)[P_j - t_j(x)] \equiv$ there is a continuous series of places including P_i and P_j and a continuous series of times including t_i and t_j such that for any time t on the time series between t_i and t_j there is a place P of the place series between (series-between, not spatially between) P_i and P_j such that there is an individual at P at t; and for any two places P_m and P_n series-between P_i and P_j, if P_m is series-prior to P_n then there is a time t_m and a time t_n both between t_i and t_j such that t_m is earlier than t_n and there is an individual at P_m at t_m and there is an individual at P_n at t_n.[9]

The symbolic notation occurring at the beginning of Wilson's account is to be read "The x which is at spatial position P_i at time t_i is identical with the x which is at spatial position P_j at time t_j."

The idea is clearly that objects picked out as existing at different times will be identical if and only if there is a continuous series of places connecting those objects, at the times at which they are picked out, and if those places are occupied, at successive times, by material objects.

Wilson recognizes that there is a problem about interpreting the term "place." It would be objectionable, he rightly notes, to say "that the place of an object is the place of its center of gravity." For "if one object is enclosed in another, their centers of gravity might coincide and we should have two objects occupying the same 'place' at the same time. . . ."[10] So places cannot be geometric points. Rather, places must be volumes of space. But while we may be clear about what it is to have a continuous series of spatial points, what it is to have a continuous series of volumes is not equally clear. Particularly, it is not clear that this kind of continuous series is logically one-one. It will not do to say, as Wilson wants to say, that in difficult cases, e.g., that of the fission of an amoeba, we must resort to an "*ad hoc* decision and not a matter of principle. . . ."[11] If spatio-temporal continuity is not a logically one-one relation then it cannot, in any case, be a logically sufficient condition for identity. And the problem is not just with difficult fission cases. On Wilson's account, is or is not a material object spatio-temporally continuous with every piece of itself from which it becomes spatially separated? Is a piece of toast spatio-temporally continuous with every crumb which flakes off that piece of toast? Wilson provides no answers to these questions. We must, accordingly, look for some fuller account of the notion of a continuous series of places, in the sense of volumes.

Richard Swinburne holds that for material objects M and M' to be identical it is necessary and sufficient that the "properties of M and M' are, apart from the properties of spatial relation, approximately similar" (i.e., that "there be some natural description under which they are approximately similar objects"), and that they be spatio-temporally continuous. What is meant by "M and M' are spatio-temporally continuous," where M is a material object picked out as existing at time t, and M' a material object picked out as existing at time t', is, according to Swinburne.

> . . . that there was a material object M'' approximately similar . . . to both M and M' at every temporal instant t'' between t and t', such that each M'' at each t'' occupies a place contiguous with the place occupied by the M'' at the prior and succeeding instants, however precisely temporal instants are identified, the series beginning with M at t and ending with M' at t'. Most parts of the M'' of equal volume at each t'' must also occupy places contiguous with a place occupied by a part of the M'' at the prior and succeeding instants.[12]

Note first that Swinburne's account appeals to the notion of approximate similarity. Presumably Swinburne includes this requirement of approximate similarity to escape having to say radically different kinds of things, e.g., Lot's wife and the pillar of salt she becomes are identical simply because they bear certain spatial and temporal relations to one-another. This restriction has, of course, the somewhat counter intuitive result of making it impossible for very dissimilar things, such as Lot's wife and the pillar of salt she becomes, to be spatio-temporally continuous, at least in Swinburne's technical sense of "spatio-temporal continuity."

A much more serious objection to Swinburne's account is that it at least appears to make spatio-temporal continuity, and as a result identity, non-transitive relations. Consider a case where an object, A, picked out at time t, is related to an object, B, picked out at t', in the manner required by the above account. Then, on the above account, A and B are spatio-temporally continuous and hence numerically identical. But B might also be related to an object, C, picked out at t'', where t'' is later than t', in the way Swinburne's account requires. Then B is spatio-temporally continuous, and identical, with C. However, A might be neither spatio-temporally continuous, nor identical, with C, for the "natural description" under which A and B "are approximately similar objects" might not be the natural description under which B and C are approximately similar objects. (An example of this sort of case would be where Lot's wife slowly turns into a pillar of salt. At t we have a woman, at t'' a pillar of salt, and at t', between t and t'', something which is approximately similar to both a woman and a pillar of salt, though in different ways.) While it might be possible to argue that spatio-temporal continuity is not a transitive relation, no argument could show the same to be true of the identity relation.

---

[9] Wilson, *op. cit.*, p. 596.

[10] Wilson, *ibid.*, p. 597.

[11] Wilson, *ibid.*, p. 597.

[12] Richard Swinburne, *Space and Time* (New York, 1968), p. 23.

Swinburne's attempt to explicate continuity in terms of a series of contiguous places is also objectionable. Contiguity and continuity are very different notions. A continuous series of places, where places are taken to be volumes of space, would surely have to be a series of overlapping volumes of space, not just a series of entirely separate but contiguous volumes of space.

IV

It should by now be apparent that any adequate account of spatio-temporal continuity must (1) make use of some clearly defined notion of a volume of space, and (2) appeal to a clearly explicated notion of a continuous series of such volumes. Taking up the second of the above requirements first, it may be of help to begin by trying to get a clear account of what it is for *spatial points* to be spatio-temporally continuous. I propose the following account:

> A spatial point, p, at time t, is spatio-temporally continuous with a spatial point, p', at time t', if and only if: (1) There is a continuous function, f, whose domain is the set T, which consists of all the temporal points t through t', inclusive, and whose range is a set S of spatial points. (2) $f(t) = p$ and $f(t') = p'$.

We now need an account of what it is for a function of the sort appealed to in (1) above to be continuous:

> A function, f, whose domain, T, is the set of temporal points t through t', inclusive, and whose range is a set S of spatial points, is a continuous function if and only if f is continuous at any arbitrary member of T, t_1. And f is continuous at t_1 if and only if, given any distance d, $d > 0$, there is a temporal interval, $\Delta t > 0$, such that for every $t_0 \in T$, where $|t_1 - t_0| < \Delta t$ (the absolute value of t_1 minus t_0 is less than Δt), $|f(t_1) - f(t_0)| < d$ (the distance between the spatial points $f(t_1)$ and $f(t_0)$ is less than d, no matter how small d is).[13]

The idea behind the above account is, of course, that any two spatial points, with temporal designations, are continuous just in case there is a continuous "line" through space and time which connects those points. A line is continuous just in case between any two distinct points on that line there are a non-denumerable infinity of points, points which form a series having the cardinality of the continuum.[14]

The task now before us is that of modeling an account, of what it is for material objects to be spatio-temporally continuous, on the foregoing account of the spatio-temporal continuity of spatial points. In so doing it will be useful to introduce the technical notion of a chunk of space. Chunks of space are specific volumes of space, with specific locations. The easiest way to refer to a chunk of space is to refer to it as the segment of space which is in fact occupied by such-and-such a material object, or as the segment of space which would be occupied by such-and-such a material object were that object in such-and-such a spatial position. And the chunk of space a material object occupies is that segment of space which is enclosed by its surfaces. Thus, on the usage I am suggesting, we can talk of the chunk of space now occupied by my television set, and of the chunk of space which would be occupied by my television set were that set sitting on the right hand corner of my desk, with its edges lined up with the edges of my desk.

---

[13] I am indebted to Mrs. Chia-Hui S. Kuo for suggesting the above accounts to me, and for helping me see how they can be modified to fit talk of spatio-temporal continuity of material objects.

[14] The relation sketched by the above accounts is, of course, not logically one-one, since there are an infinite number of continuous "lines" connecting any two points picked out at different times. This feature will not carry over into the forthcoming account of spatio-temporal continuity of material objects.

The stage is now set for the following account of what it is for material objects picked out at different times to be spatio-temporally continuous. Call this account "MOC," for "material object continuity."

> *MOC* A material object, x, at t, is spatio-temporally continuous with a material object, y, at t', if and only if: (1) There is a continuous function, f, whose domain, T, is the set of temporal points t through t', inclusive, and whose range, R, is a set of chunks of space. (2) $f(t)=s$ and $f(t')=s'$, where s is the chunk of space occupied by x at t, and s' is the chunk of space occupied by y at t'. (3) For any $t_1 \epsilon T, f(t_1)=s_1$, and s_1 is occupied, at t_1, by some material object.

We also need an account of what it is for a function of the above sort to be continuous:

> A function, f, whose domain, T, is the set of temporal points t through t', inclusive, and whose range is a set R of chunks of space, is a continuous function if and only if f is continuous at any arbitrary member of T, t_1. And f is continuous at t_1 if and only if, give any volume $v > 0$, there is a temporal interval, $\Delta t > 0$, such that for every t_0, where $|t_1 - t_0| < \Delta t$, the volume of the union of the non-intersecting segments of $f(t_1)$ and $f(t_0)$ is less than v.

An example may be helpful here. Consider a balloon, B, first in its inflated state, at t, and later in its deflated state, at t'. As the air escapes from B it shrinks and occupies successively smaller chunks of space. But the change in volume, and in position, is continuous. That is, if we consider the chunk of space occupied by B at any arbitrary time, t_1, where $t \leqslant t_1; \leqslant t'$, and give ourselves any volume $v > 0$, *MOC* assures us that there is a t_0, $t \leqslant t_0 \leqslant t'$, indeed infinitely many, such that the volume of the non-overlapping segments of the chunk of space B occupies at t_1 and the chunk of space B occupies at t_0 is less than v, no matter how small v is. Such at least is the intended purport of *MOC*.

If we make the somewhat dubious assumption that no two material objects can occupy exactly the same chunk of space at the same time it follows that *MOC* does define a logically one-one relation.[15] For if A at t is spatio-temporally continuous with B at t' there can be no material object, C at t', which is also spatio-temporally continuous with A but occupies a different chunk of space from that occupied by B. One might think B and C could occupy different chunks of space at t', and both be spatio-temporally continuous with A, if A underwent fission at sometime between t and t'. But this is not so. For if A undergoes fission—say A is an amoeba—then, on *MOC*, neither of the results of the fission will be spatio-temporally continuous with A, because at the moment of fission, i.e., at the moment spatial separation occurs, a discontinuous change in the volumes of the individuals involved will also occur.[16]

MOC succeeds in ruling out cases of fission as cases of spatio-temporal continuity, and thus in defining a logically one-one relation, by requiring that all changes in volume and position be continuous. Unfortunately this tight requirement also makes *MOC* singularly unuseful as an account of spatio-temporal continuity. This is so simply because as a matter of fact very few, if any, of the material objects with which we are familiar persist for any significant periods of time without undergoing discontinuous changes in volume. It is simply a matter of fact that material objects are constantly absorbing and emitting electrons and other subatomic particles, or at least having atoms and molecules scraped off their surfaces. (Most observable qualitative changes, such as that brought about by painting an object, involve adding or subtracting molecules, e.g., the molecules composing the paint.) And the change in volume brought about by the gain or loss of molecules, or atoms, or sub-atomic particles, is always discrete rather than continuous.

---

[15] This assumption is dubious in that it has recently been argued, by David Wiggins, that different kinds of things, e.g., a tree and a certain aggregate of molecules, can occupy exactly the same place at the same time. But it might be replied that Wiggins has not shown that distinct *material objects* can occupy the same place at the same time, for an aggregate of molecules is not a material object, though it may be an aggregate of material objects. See David Wiggins, "On Being in the Same Place at the Same Time," *The Philosophical Review*, vol. 77 (1968), pp. 90–95. David Sanford has gone farther than Wiggins and argued that two material objects of the same kind can occupy the same place at the same time, see his article "Locke, Leibniz, and Wiggins on Being in the Same Place at the Same Time," *The Philosophical Review*, vol. 79 (1970), pp. 75–82.

[16] Some philosophers prefer talk of world lines to talk of spatio-temporal continuity, and hence prefer to talk of the splitting of world lines rather than of cases of fission. But as far as I can see the "splitting" of a "world line" just is a case of the fission of a material object. And since fission does not preserve spatio-temporal continuity on *MOC*, neither does the splitting of a world line. This is because, again, there is a dramatic discontinuous change in volume when the fission or splitting occurs. The resultant objects, or temporal parts of world lines, must be markedly less in volume than the parent object, or temporal part of a world line, just before the fission or splitting occurs.

That such changes in volume are not continuous, as continuity is defined by *MOC*, can be shown by considering what happens when a leg is amputated. (The loss of a leg is different only in degree, not in kind, from the loss of a single atom, etc.) Let Edwin be the prospective amputee, t a point in time shortly before the surgeon starts to operate, and t'' a point some time after the surgeon has finished. Edwin has two legs at t, one leg at t''. At every point in time between t and t'' Edwin either has two legs, or he has one leg. (When the surgeon is half way through with his cut Edwin does not have one and a half legs; he has two legs, one of which is cut half way through.)

Consider the interval of all points equal to and later than t at which Edwin has two legs. Call this interval "I_1." Call the interval consisting of all points equal to or earlier than t'' at which Edwin has only one leg "I_2." I_1 clearly has an upper limit. This limit, call it "t'," must either be in I_1 or in I_2. Assume Edwin at t is spatio-temporally continuous with Edwin at t'', and that t' is in I_1. Then there is a continuous function, f, which maps the points t through t'' onto chunks of space. And f must be continuous at every point in time between t and t'', inclusive. But in fact f cannot be continuous at t'. For if we take as our v any volume which is greater than zero but less than the volume of Edwin's amputated leg, there will be no interval, $\Delta t > 0$, such that for every t_0 where $|t' - t_0| < \Delta t$ and $t \leqslant t_0 \leqslant t''$, the union of the non-intersecting segments of $f(t')$ and $f(t_0)$ is less than v. In particular, there will be no t_0 later than t' where this is the case. So f cannot be continuous; Edwin before the amputation cannot be spatio-temporally continuous with Edwin after the amputation.

But perhaps t' is in I_2 rather than in I_1. (In which case I_2 is a closed interval.) Again the continuous function f required by *MOC* will not be available, because no such function can be continuous at t'. That is, at t' there will be no interval Δt of the requisite sort. In particular, where v is less than the volume of Edwin's amputated leg there will be no $t_0 < t'$ such that the union of the non-intersecting segments of $f(t')$ and $f(t_0)$ is less than v. So again Edwin before the amputation cannot be spatio-temporally continuous, in the sense of *MOC*, with Edwin after the amputation. A parallel line of reasoning will lead to the conclusion that the loss of a single hair, or flake of skin, or molecule or atom or sub-atomic particle produces a discontinuous change in the volume of the object in question.

Hence though *MOC* defines a logically one-one relation, and thus might be a logically sufficient condition for the identity of material objects picked out at different times, it is at most a very peculiar sufficient condition for identity, in that it is one which is virtually never satisfied. And if *MOC* defines the only acceptable logically one-one relation we can find, it follows that accepting spatio-temporal continuity as a logically necessary condition for identity will necessitate our holding that material objects picked out at different times are virtually never identical. And this seems to be too high a price to pay for accepting the view that there are specific logically necessary and sufficient conditions for the identity of specific kind of things.

It might now be suggested that we can loosen the requirements of *MOC* so as to make it possible for the kind of material objects with which we are familiar to stand to one-another in the relation of spatio-temporal continuity. Perhaps we can allow spatio-temporally continuous objects to exhibit limited discontinuous changes in volume. We can do this, while still preserving the logically one-one aspect of *MOC*, simply by altering our restrictions on the volume v, appealed to in *MOC*. Suppose we make the second sentence of the explication of what it is for a function to be continuous, which follows *MOC*, read "And f is continuous at t_1 if and only if, given any volume $v \geqslant \frac{1}{2}[f(t_1)]$, there is a temporal interval $\Delta t > 0$, . . ." This revised requirement allows discontinuous changes in volume, so long as each discontinuous change increases or decreases the volume of the object in question by less than 50 per cent. Yet the one-one nature of *MOC* is retained, for in fission cases—which most seriously threaten the one-one status—neither offspring will be spatio-temporally continuous with the parent, assuming neither offspring is more than half as large as the parent. And if one offspring is more than half as large as the parent, then it, but not its sibling, will be spatio-temporally continuous with the parent.

So defined spatio-temporal continuity cannot be a logically necessary condition for identity, because some things undergo discontinuous decreases in volume which are greater than that allowed by the above revision of *MOC*, without ceasing to exist, e.g., plants which are pruned in the fall. Moreover, as revised above *MOC* does not seem to be a plausible sufficient condition for identity. For on the revised view a phonograph record which is broken into two pieces of unequal size will be spatio-temporally continuous with the larger of those pieces. But it seems false to conclude from this that the phonograph record is identical with the larger of those pieces. When we break a record we are left with pieces of a record, no one of which is itself a phonograph record.

But perhaps the revised version of *MOC* can be the basis for an acceptable logically sufficient, but not necessary, condition for identity. The problem in the record case is that neither of the products of breaking a record seems to be a record. A whole record and parts of a record are different kinds of things. We can perhaps take account of this difference by adopting the Geach and Anscombe like view that if x and y are identical, then "x and y are the same *F*" will be true for some replacement of 'F' by a substantival term. The general requirement might then be that x and y, material objects picked out at different times, are identical if they are spatio-temporally continuous, in the sense of our revised *MOC*, and "both" are correctly characterized by some one substantival term.

The foregoing view would, of course, have to be refined. Criteria for what terms are to count as substantival terms would have to be developed, and so on. But what is more important for our present purposes is that the envisioned requirement would be, at most, a logically sufficient condition for identity, not a logically necessary condition. For things which are not spatio-temporally continuous in the sense of our revised version of *MOC*, e.g., a raspberry bush before an extensive pruning and a raspberry bush after such a pruning, can—or so we ordinarily claim—be identical. Now if we accept the view that there is even one logically sufficient condition for identity, e.g., one based on our revised *MOC*, then we are surely committed to finding one or more sufficient conditions for identity which will cover the remaining cases of the identity of material objects picked out as existing at different times—cases where discontinuous increases and decreases in volume of more than 50 per cent do occur. These additional sufficient conditions must, of course, be based on logically one-one relations. The problem is just that no such relations seem available. The requirement that x and y "both" be correctly characterized by at least one substantival term, in the absence of the additional requirement of spatio-temporal continuity, will not, of course, do. Jane's desk and Mary's desk are both correctly characterized by the substantival term "desk," but they need not be the same desk.

There are, moreover, some serious difficulties with *MOC*, both as originally formulated and as revised above. In the first place the assumption that no two material objects can occupy exactly the same chunk of space at the same time may prove to be false, as was noted above. If it is false, then *MOC* does not define a logically one-one relation. Secondly, *MOC* assumes that the notion of a chunk of space can be rigorously defined. This just seems unlikely, despite the attempt made above to explicate that notion by talking of the surfaces enclosing material objects. The trouble is with the notion of the surface of an object. It takes only a passing acquaintance with modern physics to realize that the boundaries of material objects, their surfaces, cannot be specified with more than a rough degree of accuracy, for the atomic components of material objects are in constant motion. When we become concerned with very small distances the amorphous, ever changing, boundaries of a fog bank provide a better model of the surfaces of material objects than do the cleanly cut idealizations of solid geometry.[17]

Perhaps *MOC* is radically mistaken, and some other account of spatio-temporal continuity can be given, an account which makes that relation, or some relation built on it, logically one-one, and does so in such a way that the kinds of material objects with which we are familiar frequently do stand to one-another in this relation. But this just seems unlikely. Indeed it now seems that talk of the spatio-temporal continuity of material objects is a loose and popular way of talking, a way of talking which does not allow of rigorous refinement. If this is the case, then spatio-temporal continuity is neither a logically necessary nor a logically sufficient condition for the identity of material objects picked out as existing at different times.

My argument for the thesis that we neither have, nor are likely to find, any interesting logically necessary and/or sufficient conditions for the identity of material objects picked out as existing at different times is now complete. And the conclusion that we shall have to be content with the ferreting out of the general, but violable, principles which we do use in deciding, in particular cases, what pairs of objects picked out as existing at different times do satisfy the requirement of schema (1), are identical, is now warranted. The ferreting out of these general principles may or may not be a philosophically interesting project. And the results reached above may or may not bear on the search for specific logically necessary and/or sufficient conditions for the identity of kinds of things other than material objects picked out as existing at different times.[18]

---

[17] I owe this point to Robert Coburn.

Temple University

Received January 25, 1971

---

[18] I wish to thank my colleagues, Gerald Vision and David Walker, and a referee of this journal, for their helpful comments on an earlier version of this paper.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 2, April 1972

VII. THE LOGIC OF PASCAL'S WAGER

IAN HACKING

PASCAL'S wager is the name of some game-theoretic considerations that concern belief in God.[1] I shall show that Pascal briefly presents three distinct arguments, each with different premisses. In each case, the conclusion does follow from the premisses. In each case, there is at least one unacceptable premiss. I am here concerned only with the logic of the wager, that is, with the question of whether the conclusion does follow from the premisses.

My aims are twofold. First, Pascal's wager has been extensively studied, especially in recent times when the existential and non-deductive characteristics of the argument have seemed more important to moralists and theologians than traditional, more deductive, ways to God. Despite such studies, the logic of Pascal's three arguments has never been set out. Moralists tend not to study decision theory. Pascal invented it. This paper may help the moralists who read his thoughts to understand his invention.[2]

Secondly, I have the historical aim, divorced from morality, of stating a bit of the history of probability. Pascal's arguments show that it was possible to have a deep understanding of what we now call decision theory, even in the 17th century. Possible, but not easy: Pascal's distinctions escaped most of his readers. Yet despite misunderstanding, Pascal's wager was a decisive turning point in probability theory. Pascal is celebrated for the probability correspondence with Fermat, but that is not his most profound contribution to probability.[3] Pascal's thought showed that the mathematics of games of chance had quite general applications. He made it possible for *the doctrine of chances* (a theory involving "objective" frequencies, whose primitive concept is expectation, or average winnings) to become *the art of conjecture* (a "subjective" theory about degrees of confidence).[4]

I. DECISION THEORY

Decision theory is the theory of deciding what to do when it is uncertain what will happen. *Given* an exhaustive list of possible hypotheses about the way the world is, the observations or experimental data relevant to these hypotheses, together with an inventory of possible decisions, and the various utilities of making these decisions in various possible states of the world: *determine* the best decision.

A special case of this problem occurs when no experiments are made. In the thought that concerns us, Pascal deliberately "ties his hands" and refuses to look at any observations or experimental data bearing on the existence of a Christian God. He is writing for the man who will not countenance miracles, or the doctors of the church, or the witness of the faithful. So we may restrict our attention to the logic of decision when there is no experimental data.

---

[1] Blaise Pascal, *Pensées sur la religion et sur quelque autres sujets*, the first edition, *édition de Port-Royal*, was in 1670. The standard text is that of Louis Lafuma (Paris, 1951); or *Oeuvres Complètes*, ed. by Lafuma (Paris, 1963). Pascal's wager is the fragment headed *Infini-rien*, which is no. 418 of Lafuma. Many older editions and translations use Leon Brunschvicg's numeration, where *Infini-rien* is 233.

[2] Pascal did not invent the notion that belief in God is better than disbelief because belief brings salvation if there is a God, and costs little if there is no God. He invented the concepts of decision theory which create valid argument forms out of this earlier idea. For precursors of Pascal (who did not understand Pascal's logic) see Georges Brunet, *Le Pari de Pascal* (Paris, 1956), p. 62.

[3] The 1654 letters to Fermat are in the *Oeuvres Complètes*, *op. cit.*, pp. 43–49. There are several fabulous anecdotes about the origin of this correspondence in general circulation; for a good antidote, consult Oystein Ore, "Pascal and the Invention of Probability Theory," *American Mathematical Monthly*, vol. 67 (1960), pp. 409–419. For a complete historical background, see Jean Mesnard, *Pascal et les Roannez* (Bruges, 1965).

[4] The art of conjecture and what used to be called the subjective theory of probability begins with Jacques Bernoulli, *Ars conjectandi* (Basel, 1713). Carnap has argued that interpretations of probability as objective long run frequencies commence only in the 19th century (*Logical Foundations of Probability* [Chicago, 1950], §42). In fact, as we remark in passing in Section 7 below, the frequency-style interpretation is the oldest, although Pascal began to change all that. Proofs of this opinion will be published elsewhere. The doctrine of chances is an old name for the study of physical properties of games of chance, although the most celebrated work of that title, by Abraham de Moivre, was published five years after the appearance of Bernoulli's masterpiece.

Among the valid argument forms investigated by decision theory, three were, I shall show, apprehended by Pascal. I call them "valid" in the sense now favored by logicians. A valid argument form is one in which the conclusion follows from the premisses. Colloquially, of course, a valid argument is one that is pertinent and persuasive. Nowhere in this paper shall I imply that Pascal's arguments are persuasive. But this is not because they are invalid (in the logician's sense) but because the premisses of the arguments are, at best, debatable. Here are the three argument forms that will concern us:

(a) *Dominance.* The simplest special case occurs when one course of action is better no matter what the world is like. Schematically, suppose that we have some exhaustive set of possible states of affairs: we label the states S_1, S_2. . . . Suppose that in some state S_i the utility U_{i1} of performing act A_1 is greater than the utility U_{i2} of performing act A_2. In no other state of affairs is the utility of performing A_1 less than A_2. Then, under no circumstances could A_2 have happier consequences than A_1, and under some circumstances A_1 could be better than A_2. A_1 is said to dominate A_2. If one act dominates all others the solution to our decision problem is, "*perform the dominant act.*"

(b) *Expectation.* The argument from dominance does not consider how likely are various states of affairs. Even if dominating A_1 is better only in a very unlikely state of affairs, then, since A_1 can never fare worse than any other act, it is worth while performing A_1. But suppose no action dominates, although we do think we know which states of affairs are more likely than others. Suppose we can assess the probability of each state of affairs. Then (no matter what one means by "probability") one argues as follows. We have assigned a probability p_i to each possible state of affairs S_i in some exhaustive set. Let U_{ij} stand for the utility of doing A_j if S_i actually obtains. The *expected value*, or *expectation*, of A_j is the average[5] value of doing A_j: namely $\sum_i p_i U_{ij}$. An argument from *expectation* concludes with the advice, "*Perform an act with highest expectation.*"

(c) *Dominating expectation.* It may happen that we do not know, or cannot agree on, the probabilities of the various states of affairs. We can at best agree on a set of probability assignments to the states S_i. For example, suppose we agree that the coin is biased toward heads, but disagree how great is the bias; at least we agree that the probability of 'next toss gives heads" exceeds 1/2. If in some admissible probability assignment, the expectation of A_1 exceeds that of any other act, while in no admissible assignment is the expectation of A_1 less than that of any other act, then A_1 has dominating expectation. The argument from dominating expectation concludes, "*Perform an act of dominating expectation.*"[6]

The three argument schemes are mutually consistent. If one act does dominate the rest, then it will be recommended by all three arguments. If there is no dominating act, but an act of highest expectation, that act will also be the act of dominating expectation. The argument from dominance is the rarest, most special case. The argument from dominating expectation is more widely applicable.

Pascal's procedure in the thought "infini-rien" is to offer (a) an argument from dominance. Then, if its premisses be rejected, to offer (b) an argument from expectation. Then, if the second lot of premisses be rejected, to offer (c) an argument from dominating expectation.

II. PASCAL'S THOUGHT

Immense scholarship has been lavished on *infini-rien*: it consists of two pieces of paper covered on both sides by handwriting going in all directions,

---

[5] Someone unacquainted with probability jargon may wonder why the expectation is called an "average." This innocent question probes quite deep. The answer depends on what one means by probability. If probability is long run frequency, then the expectation is quite literally the average profit. For example, if act A_j is, "bet on heads" and a coin falls heads 2/3 of the tosses, and if the pay-off for a correct bet is \$5, while the fine for an incorrect one is \$4, the average profit on A_j is

$$\tfrac{2}{3}(\$5) + \tfrac{1}{3}(-\$4) = \$2.$$

If probabilities are personal degrees of confidence, the expectation is not, strictly speaking, an average. The validity of this argument form, under this interpretation, may be derived from a joint analysis of personal probability and utility, as in F. P. Ramsey, "Truth and Probability," in *The Foundations of Mathematics* (London, 1931), pp. 156–198. If probabilities are long run frequencies, and one is repeatedly wagering, then maximizing expectation is justified as bringing the best average profits. If one wagers only once, however, the validity of this pattern of argument may be questioned.

[6] If probabilities are my personal degrees of confidence, then I may assign unique probabilities to each state S_i, and get along with argument form (b). Even so I may need (c) in arguing with others. Like Pascal, I may say, "our probability assessments differ, but at least we agree the probabilities lie in this range, and there is still an inference from dominating expectation."

full of erasures, corrections, insertions, and afterthoughts.[7] There is endless speculation about when and how Pascal wrote these two pages; every blot of ink or raindrops on them has been given minute analysis. We do not presume to trespass on this critical work, and take for granted Lafuma's text.

"Infini-rien" has also attracted a great deal of philosophical scholarship from moralists and theologians alike. But although they have been aware of the rough lines of Pascal's argument, their own analyses have been defective because they have not noticed that Pascal has three distinct arguments. First he offers an argument from dominance. But dominance occurs only if a certain premiss is admitted. If it is rejected, a new tack must be pursued. We get an argument from expectation. Yet this requires an implausible assignment of probabilities. So the argument is changed again to an argument from dominating expectation. Not all the links forged by Pascal are transparent, but it is remarkable how these two scratched sheets that constitute "Infini-rien" conform to this analysis. Here are the details.

The argument is directed, as the Port Royal editors say from the start, at the sort of person who "not being convinced of the proofs of religion, and still less by the arguments of atheists, remains suspended between a state of faith and one of unbelief."[8] This assertion is extremely important. A decision problem requires an exhaustive partition of possibilities. It is taken as a premiss of the argument that either there is no God, or else there is a God whose characteristics are correctly reported by the Church. Mohammedanism, for example, is not admitted as a possibility. It is a corollary, that Pascal's argument is good for any decision problem with the same formal structure. "An Imam could reason just as well this way," as Diderot remarked.[9]

Pascal's partition may be out of place today. But this is one thought from a book of thoughts; the other thoughts contain other reasons bearing on this partition. There are also other arguments directed at other special sorts of person; for example, the arguments directed at Orthodox Jews whose partition, of course, is not at all like that of the Parisian man about town.

"God is, or he is not." This is the way Pascal abbreviates the partition of the Parisian. "Which way should we incline? Reason cannot answer." That is, there is no valid proof or disproof of God's existence. Instead we adopt the following model:

> A game is on at the other end of an infinite distance, and heads or tails is going to turn up. Which way will you bet?

The model is then reinforced. When reason cannot answer, a sensible man can say that he will not play the game. But in our case, by the mere fact of living, we are engaged in play. We either believe in God, or we do not.

A decision problem requires a partition of possible actions. As Pascal sees it, you either act with complete indifference to God, or you act in such a way that you will, in due course, believe in his existence and his edicts. There is no cant to Pascal. He accepts as a piece of human nature that belief is catching: if you go along with pious people, give up bad habits, follow a life of "holy water and sacraments" intended to "stupefy one" into belief, you will become a believer. Pascal is speaking to a man unsure whether to follow this path, or whether to be indifferent to the morality of the church. The two possible acts are *not*, "Believe in God" and "Do not believe." One cannot decide to believe in God. One can decide to act so that one will very probably come to believe in God. Pascal calls that the wager that God is. To wager that He is not is to stop bothering about such things.

III. The Argument from Dominance

The decision problem is constituted by two possible states of the world, and two possible courses of action. If God is not, both courses of action are pretty much on a par. You will live your life and have no bad effects either way from supernatural intervention. But if God exists, then wagering that there is no God brings damnation. Wagering that God exists can bring salvation. Salvation is better than damnation. Hence the wager, "God is," dominates the wager, "he is not." The decision problem is solved.

The argument is valid. The premisses are dubious, if not patently false. Few non-believers now can suppose that Pascal's partition exhausts the possibilities. If we allow just one further alternative, namely the thesis of some fundamentalist sects, that Jehovah damns all who toy with "holy water and sacraments," then the Catholic strategy no longer dominates. That is, there is one possible state of affairs, in which the Catholic strategy does not have the best pay-off.

---

[7] For a photograph, see, for example, Henri Gouhier, *Blaise Pascal, commentaires* (Paris, 1966), or G. Brunet, *op. cit.*, n. 2.

[8] *E.g.*, in the 6th (unnumbered) page of the preface to the edition of Amsterdam, 1677.

[9] Denis Diderot, *Pensées philosophiques*, LIX, *Oeuvres*, ed. by J. Assézat (Paris, 1875–1877), vol. I, p. 167.

It is a question for the historian whether Pascal's argument from dominance could ever have been effective. Perhaps some Parisians three centuries ago did believe, say, on the evidence of alleged revelation, that if there is any supernatural truth, it is that truth professed by the Church. Such people may have found the argument persuasive.

From the very first, a fallacious correction has sometimes been urged. M. J. Orcibal has found the following contemporary note by Daniel Huet: "this reasoning suits all religions; that which proves too much proves nothing. It proves only the necessity of having some religion, but not the Christian religion."[10] This is a mistake; unless one has independent grounds for excluding the proposition of an eccentric religion, that all and only religious people are damnable. This point has been nicely made recently by James Cargile.[11]

One premiss needed for Pascal's argument is that faith is catching. His interlocutor does protest, that perhaps a treatment of holy water and sacraments is not going to work. The theme is not well developed.[12] Also, many of us will share William James's suspicion that a person who becomes a believer for the reasons urged by Pascal is not going to get the pay-offs he hopes for.[13] Although all these questions are crucial to the merit of the argument as a piece of apologetics, they are irrelevant to its logical validity. The conclusion does follow from the premisses.

IV. The Argument from Expectation

Pascal is more worried by another premiss, as yet hardly stated. We assumed that if God is not, then either course of action has roughly equal utility. But this is untrue. The libertine is giving up something if he chooses to adopt a pious form of life. He likes sin. If God is not, the worldly life is preferable to the cloistered one. The wager, "God is," does not dominate, for there is one circumstance (God is not) in which accepting this wager, i.e., adopting the pious life, has lower utility than the other wager. When dominance is challenged, we require an argument from probabilities.

The transition is swift, perhaps too swift. Having heard the argument from dominance, the interlocutor says, "very well. But suppose I am asked to stake too much?" This remark is obscure unless seen in the light of the present analysis. The interlocutor is protesting that he has to give up something to follow the pious life. Perhaps he stakes too much, that is, perhaps there is a significant difference between the utility of the two actions, under the hypothesis that there is no God. So Pascal has to introduce the policy of maximizing expectation. If the chances of heads and tails are equal, and there are equal pay-offs for either outcome, it is a matter of indifference whether we bet on heads. But if heads pays twice as much as tails, then clearly we bet on heads. In the agnostic's existential situation, the optimal pay-off, if there is no God, is a worldly life. The optimal pay-off, if there is a God, is salvation, of incomparably greater value. Hence, if there is an equal chance of God's existence or non-existence, the expectation of choosing the pious life exceeds that of choosing the worldly one. The argument from expectation concludes, act so that you will come to believe in God.

V. The Argument from Dominating Expectation

The argument from expectation can hardly be maintained. Although it is valid, it requires a monstrous premiss of equal chance. We have no good reason for picking 1/2 as the chance of God's existence. This argument can work only for people who are, in the strongest sense, exactly as unsure whether God exists, as they are unsure whether he does not exist. Against all other agnostics, another argument is called for.

The argument from expectation with an equal probability distribution requires only that salvation, if God is, is more valuable than sinful pleasures, when there is no God. But salvation is infinitely preferable to the joys of the worldly life. No matter how great may be the daily pleasures of the libertine, they are finite. Salvation, according to Pascal, is infinitely blessed.

---

[10] M. J. Orcibal, "Le fragment infini-rien et ses sources," p. 183, n. 89 in *Blaise Pascal, l'Homme et l'Oeuvre*, Cahiers de Royaumont (Paris, 1956).

[11] James Cargile, "Pascal's wager," *Philosophy*, vol. 41 (1966), pp. 250–257.

[12] One may pedantically urge a partition into three states of affairs: God is not, God is, and one will come to believe by being pious; and, God is, but one will not come to believe by being pious. The pious strategy still dominates, under the initial assignment of utilities favored by Pascal.

[13] William James, "The Will to Believe" in *The Will to Believe and other Essays in Popular Philosophy* (London, 1897).

Moreover, although we have no idea of the chance that God exists, it is not zero. Otherwise there would be no problem. There is a finite positive chance that God exists. No matter what this finite chance is—no matter how small—the expectation of the pious strategy with infinite reward exceeds that of the worldly one. Hence the pious strategy must be followed. This is an argument from dominating expectation.

VI. Summary

All three arguments are valid. None are convincing. All rely on dubious premisses. The arguments are worthless as apologetics today, for no present agnostic who understood the arguments would ever be moved to accept all the premisses.

The most dubious premiss is the partition, with its concomitant assignment of utilities. God is (and belief in Him brings salvation), or, God is not (and non-believers who have heard the Word are damned). It is no criticism of Pascal that he assumes this partition: he is directing his argument at his fellows who accept it. He knew this, and so did his editors. All three arguments assume this partition and the assigned utilities.

Many questions of morality and theology may be adduced against the wager. What is striking is that Pascal raised the logically relevant questions. There is a logically special feature of the argument from dominance, which is challenged if the utility of a libertine life exceeds that of a pious one, in the event that there is no God. To allow for this difficulty, we move to the second argument form, employing probabilities. An equal distribution of probabilities has some appeal, but that special logical feature is in turn abandoned as we move to the third argument form. Pascal's readers, unfortunately, did not possess his subtlety. They got some whiff of the argument from dominance and a vague bearing on probability; I set out such curiosities in Section VIII below. First a word on how Pascal conceived probability.

VII. Probability

Throughout this analysis I have spoken of probabilities. I have been neutral, for example, between statistical and inductive probabilities. Did Pascal suppose that the chance that God exists is an objective physical fact? Does His existence bear some logical relation to the evidence short of deductive implication? Or are we here concerned with personal degrees of belief?

If someone wanted to use Pascal's argument today, he would almost certainly use the degree of belief interpretation. It makes no sense to think of our world having been determined as the outcome of a trial on a chance set-up whose possible outcomes are, God is, and, God is not. But it would be wrong to infer that Pascal had adopted an interpretation of probability in terms of degrees of belief. Maybe he should have adopted it, but he did not.

Pascal's language is consistently that of the doctrine of chances. We are explicitly concerned with a game and the flipping of a coin. In the final argument, from dominating expectation, there is, he says, at least one *hasard* that God exists; there are at most finitely many *hasards* that He does not exist. This is the language used in describing a lottery, where each ticket is called an *hasard*. Pascal's thought seems to be: each agnostic is in the same position as a gambler in a game with certain constraints. In virtue of certain objective physical features of the game, the rational gambler must make a certain decision. Because the agnostic is in certain respects in the same position as that gambler, the rational agnostic must make the like decision. The only probability notion which Pascal actually uses is *hasard*, and there is no evidence that he interprets this notion in terms of degrees of belief. Certainly his ideas would lead one to that interpretation. Pascal has not gone that far.

VIII. Pascal's Readers

Port Royal editors produced the *Pensées* in 1670, eight years after Pascal's death. The immediate fame, or notoriety, of the wager is not attributable only to this edition. For part of the argument had been presented earlier in a simple and defective form on the concluding pages of the Port Royal logic.[14] There is no difficulty in detecting the origin of this idea of Arnauld, the supposed author of the final part of the *Logic*. Indeed, the origin was immediately noticed by Leibniz.[15] Arnauld learned Pascal's lesson well. He concludes with a sketch of the wager, but just before that he describes more mundane decision problems. Those chapters include the very first time (I believe) when something actually called probability is measured. Thus from the very beginning, numerical probabilities were invoked for quite well defined decision problems.

---

[14] *La Logique, ou l'Art du Penser* whose first edition appeared in 1662. The chief authors were Antoine Arnauld and Pierre Nicole. This is a manual of lessons, prepared for students at Port Royale des Champs, the Jansenist retreat that succored Pascal.

[15] In a letter to his patron, 1678. G. W. Leibniz, *Sämtliche Schriften und Briefe*, ser. I, vol. 2, p. 112.

Locke picked up Pascal's argument, like much else, from the Port Royal Logic. His presentation of it in the *Essay on Human Understanding* excited much undeserved admiration. Locke thinks that the whole argument follows for anyone who

> will allow exquisite and endless happiness to be but the possible consequence of a good life here, and the contrary state the possible reward of a bad one. . . . I have forborne to mention anything of the certainty or probability of a future state, designing here to show the wrong judgement that any one must allow he makes, upon his own principles, laid how he pleases, who prefers the short pleasures of a vicious life upon any consideration, whilst he knows, and cannot but be certain, that a future life is at least possible.[16]

Locke, it is evident, had no conception of probability logic. In his *Nouveaux Essaies* Leibniz simply passes by Locke's version of the argument. By the time he read Locke, the argument had no novelty. In a letter dated 1683, Leibniz' opinion on the merits of the wager is unequivocal. He grants that it may have some uses. Every man must put every argument to what use he can. However:

> Pascal paid attention only to moral arguments, such as he presented so well in his little posthumous book of thoughts, but he did not put much value on the metaphysical arguments which Plato and St. Thomas and other philosophers and theologians have used in proving the divine existence and immortality of the soul. In this I do not agree with him. I think that God speaks to us, not merely in sacred and civil history, but also internally, within our mind, through truths which abstract from matter, and are eternal. Even if I should confess that these arguments have not been carried to the full force of a complete demonstration, they already seem to have as much force as the moral arguments. I believe that men will gradually perfect them.[17]

Pascal's opinion was exactly the opposite. "The metaphysical proofs of God are so removed from the reasoning of mankind, and so complicated, that they have no force." This thought, though appearing in all arrangements of the *Pensées* a long way from "Infini-rien," has been shown by Georges Brunet to be one of four jottings written on the same kind of paper as that Pascal used for his wager.[18] It is inferred, with some likelihood, that it was much in his mind at the time.

Our long quotation may suggest that Leibniz was a little condescending about Pascal's style. Quite the contrary. A respected apologist named Michel Mauduit anonymously printed a version of the argument in 1677. Next year Leibniz contrasts the rhetoric of this book with that of Pascal: "Mauduit is not for people who reason with application, and who prefer naked thought as in Pascal, where it is stripped of useless words."[19]

Mauduit is not that bad: he gives a studious explication of Pascal's arguments. He states, in the preface, that the discussion in Pascal "is too short and too fast for something so important." He fills it out, and treats objections. He states the simple principle of dominance as a maxim, "used in commerce and all branches of life."[20] I have not found the maxim stated earlier, even though it is in a sense obvious. Indeed I have not found it stated, *as a fundamental maxim of decision theory*, in any subsequent discussion of Pascal. Despite such occasional insights, Mauduit's tract is singularly insubtle. He hardly realized that Pascal's version of the wager is based on the premiss that faith is catching. Pascal is free from, but Mauduit deserves, these remarks of Leibniz:

> This argument shows nothing about what one ought to believe, but only about how one ought to act. That is to say, it proves only that those who do not believe in God should act as if they did.[21]

Pascal would add: because, if they do so, they will in due course come to believe. Now that is no reason for believing. It is a cause of believing. Leibniz is a rationalist about belief. Pascal understands belief better than Leibniz.

Travesties of Pascal's argument became common. In 1699 James Craig published a version of it which relied on the testimony of witnesses.[22]

---

[16] John Locke, *Essay Concerning Human Understanding*, Bk. II, ch. 21, §72.
[17] In a letter to von Seckendorff, 1683. *Op. cit.*, ser. II, vol. 1, p. 533.
[18] Georges Brunet, *Le Pari du Pascal* (Paris, 1956), pp. 48–51.
[19] *Loc. cit.*, n. 15.
[20] Michel Mauduit, *Traitté de Religion contre les Athées, les Déistes et les nouveaux Pyrrhoniens* (Paris, 1677).
[21] G. W. Leibniz, *Die Philosophischen Schriften*, ed. by C. I. Gerhardt (Berlin, 1875–1890), vol. III, p. 621.
[22] John Craig, *Theologiae Christianae principia mathematica* (London, 1699).

Certain events are said to have happened in the Holy Land a long time ago. How reliable are they? Craig supposes that the credibility of the stories diminishes at each telling. By now the stories have worn a little thin. But because they still have finite probability, and because of the infinite payoffs for believing them if they are correct . . . and so on.

Craig's book was popular enough to be republished in 1755. Laplace amused himself by demolishing the argument, and indeed turns it on its head. The credibility of a witness is in part a function of the story being reported. When the story claims to have infinite value, the temptation to lie for personal benefit is asymptotically infinite. Hence the biblical tales do not have even finite probability, and the argument collapses. Curiously, Laplace takes this to refute Pascal.[23] But Pascal explicitly rejects any appeal to witnesses. Laplace can have read only Craig or something similar.

Whatever its value as an apologetic, Pascal's logic remained. Games were seen to serve as models of all sorts of decision under uncertainty. Voltaire was too late when he said of "Infini-rien,"

> This article seems a little indecent and puerile: the idea of a game, and of loss and gain, does not befit the gravity of the subject.[24]

Peterhouse, Cambridge

Received January 18, 1971

---

[23] P. S. de Laplace, *Essais Philosophique sur les Probabilités*, ch. XI., *Oeuvres* (Paris, 1886), p. lxxxvii.

[24] Voltaire, "Remarques sur les Pensées de M. Pascal," 1728, in *Oeuvres*, ed. by Garnier (1879), vol. XXII, pp. 32–33.

American Philosophical Quarterly
Volume 9, Number 2, April 1972

VIII. TWO TYPES OF MOTIVE EXPLANATION

STANLEY L. PAULSON

IN explaining behavior by an appeal to the agent's motive, is one thereby appealing to a disposition? Clearly not, one will reply. But why not? The reason commonly given is that one may be able to explain the agent's behavior by appealing to a motive from which the agent is not, as a matter of fact, disposed to act.[1] A very different reason, not often heard but equally appropriate, is that one may be able to explain the agent's behavior by appealing to a motive with respect to which the question of a disposition simply does not arise.

The former reason, given in discussions of the dispositional theory of mind,[2] does not call into question a tacit assumption of that theory—that all of the psychological expressions used to give motive explanations have forms which are used to ascribe dispositions. These psychological expressions, in other words, are of a single type. Given this assumption, a mistaken one I believe, it is not surprising that a second type of psychological expression has been overlooked. Expressions of this second type are also used to give motive explanations, but they have no forms which are used to ascribe dispositions.

These two types of psychological expressions yield motive explanations of two logically distinct types, and the characterization of these explanations is my concern in what follows. In the first of three sections I introduce psychological expressions of both types and show how they can be distinguished in terms of the conditions of their use in predication. In the second section I argue that the distinction between these two types of expressions is reflected, in important ways, in the motive explanations in which they are used. In the third section I provide a means of distinguishing motive explanations, of both types, from still other types of psychological explanation.

I

Speakers can readily distinguish psychological expressions which are used to give character descriptions from those which are not. For example, speakers will agree that the psychological expression "greedy" is used to give a character description in the statement "Jones is greedy," and equally, they will agree that the psychological expression "grateful" is not used to give a character description in the statement "Jones is grateful." One's statement that Jones is greedy will be true only if the expression "greedy" can be correctly predicated of Jones over a period of time. If a person behaved in a single instance as though he were greedy, one would be likely to speak of such behavior as uncharacteristic or out of character. The situation with the expression "grateful" is rather different. To be able correctly to predicate the expression "grateful" of Jones over a period of time is by no means a necessary condition of the truth of one's statement that Jones is grateful. Although it may be uncharacteristic of Jones to be grateful, that does not prevent one from saying—and saying truthfully—that Jones is grateful in a particular situation.

The point can also be made in another way. That the conjunction of statements (i) "He is greedy" and (ii) "It is uncharacteristic of him to be greedy" is strained, suggests that the expression "greedy" is typically used only of one who is characteristically greedy. In other words, it suggests that the expression "greedy" is used to ascribe a disposition. Knowing that a person is greedy, one can predict aspects of that person's behavior in a number of hypothetical situations. That the conjunction of statements (iii) "He is grateful" and (iv) "It is uncharacteristic of him to be grateful" is not strained, suggests that the use of the expression "grateful" is not determined by whether or not one is characteristically grateful. It suggests that the expression "grateful" is not used to ascribe a disposition. The distinction between characteristic and uncharacteristic simply does not arise in determining whether predicative uses of the expression "grateful" are appropriate.

---

[1] See, among others, Philippa Foot, "Free Will as Involving Determinism," *The Philosophical Review*, vol. 66 (1957), p. 446; Elizabeth Anscombe, *Intention* (Oxford, 1957), pp. 20–21; Peter Winch, *The Idea of a Social Science* (London, 1958), p. 80; A. I. Melden, *Free Action* (London, 1961), p. 12; Anthony Kenny, *Action, Emotion and Will* (London, 1963), p. 77; and A. R. Louch, *Explanation and Human Action* (Oxford, 1966), p. 96. Examples provided by Foot and others suggest, correctly, that one can give an explanation of behavior which is uncharacteristic or out of character by simply using the substantival form of an expression ("vanity") which in its adjectival form ("vain") can be used to ascribe dispositions—see, for instance, Foot's example of one who "acts out of vanity on a particular occasion although he is not in general vain."

[2] Gilbert Ryle, *The Concept of Mind* (London, 1949), chs. 4 and 5.

But why suppose, as I have, that only characteristic properties are dispositions? Knowing that Jones is grateful in a particular situation one may be able to predict aspects of Jones's behavior in that situation. So, the argument might continue, just as "greedy" may be used to ascribe a disposition, likewise "grateful" may be used to ascribe a disposition, albeit a short-term one. I have two reasons for rejecting the argument. First, predictive power, while necessary, is not sufficient for the claim that a property P is a disposition. Knowing that Jones is obese, or stupid, one may be able to predict aspects of Jones's behavior in a number of hypothetical situations; but speakers will agree that neither of these properties is a disposition of Jones's. Secondly, the limitations of prediction in a case in which one uses "grateful" to explain behavior are similar to the limitations of prediction in cases in which one uses the substantival form of "greedy" to explain behavior, including behavior which is uncharacteristic or out of character. The similarity is important, for one will not say that a person is disposed to do that which is uncharacteristic. Knowing, for example, that Jones acted out of greed in a single, exceptional situation, one may be able to predict aspects of Jones's behavior in that situation. But Jones' behavior is, by hypothesis, uncharacteristic; knowing, simply, that Jones acted out of greed in a single, exceptional situation, one could not predict aspects of Jones's behavior in other situations. Compare "grateful." From the knowledge, simply, that Jones is grateful in a particular situation, one may be able to predict aspects of Jones's behavior in that situation. But one does not thereby have a basis for predicting aspects of Jones's behavior in other situations; one lacks knowledge that Jones would be grateful in these situations.

Those psychological expressions whose adjectival forms are used to ascribe dispositions and whose adjectival and substantival forms are used to give motive explanations, I shall call *diathetic* expressions (deriving the adjective from the Greek word "*diathesis*," meaning disposition). Those psychological expressions whose adjectival and substantival forms are used to give motive explanations but which have no forms used to ascribe dispositions, I shall call *adiathetic* expressions. Lists of diathetic and adiathetic expressions would include—but would not be limited to—the following terms, given here in their substantival forms:

Diathetic	*Adiathetic*
ambition	admiration
avarice	curiosity
compassion	envy
curiosity	gratitude
greed	hatred
insecurity	jealousy
jealousy	love
revenge	pity
vanity	remorse
vindictiveness	sympathy

Several of these terms are used sometimes as diathetic and other times as adiathetic expressions, and I have included them in both lists. I shall have something to say about them below.

In what follows I use the expression "greed" (in its several forms) as a representative example of a diathetic expression, and the expression "gratitude" (likewise in its several forms) as a representative example of an adiathetic expression; and I use the factors named by these expressions (namely *greed* and *gratitude*) as representative examples of diathetic and adiathetic properties.

Can one provide a general description of the distinction marked by the speaker's use of diathetic and adiathetic expressions? By considering what I shall speak of as the *objects* associated with psychological properties one can, I shall argue, provide such a description. Of particular importance here is the nature of the connection that exists between psychological properties and their objects. Beginning with an adiathetic case, when one says that someone is grateful one is ordinarily saying that he is grateful *to someone, for something*, as in the case of Roe, who is grateful to Doe, for a favor. The person to whom Roe is grateful (namely, Doe) and that for which Roe is grateful (namely, a favor) together constitute the objects of Roe's gratitude. I shall regard the objects of Roe's gratitude in this particular instance as examples of the kinds of objects which are associated (in pairs) with gratitude generally. And one may correctly say that a person is grateful at a particular time only if there are, present to that person at that time, objects of kinds associated (in pairs) with gratitude.

Compare the situation with the diathetic expression "greedy." To say that Jones is greedy is to say that he is characteristically greedy. Not surprisingly then, it may be possible correctly to predicate the expression "greedy" of Jones at a particular time even though there is nothing for which Jones is greedy at that time. An object of a kind associated with greed, while present to the agent when he acts from greed, need not be present to the agent as a condition of his being a greedy person.

I believe the distinction marked by the speaker's use of the diathetic and adiathetic expressions "greedy" and "grateful" reflects the distinction between diathetic and adiathetic expressions generally. And I am suggesting that this distinction can be described in terms of the distinct connections existing between diathetic and adiathetic properties and their objects. An abstract statement of the distinction follows:

> If it is possible correctly to predicate a psychological expression '*P*' of a person *a*, so that the statement "*a* is *P*" is true at a particular time t_1, even though at t_1 there may be no object or objects with respect to which *a* is *P*, then '*P*' is a diathetic expression in that statement. If it is not possible correctly to predicate the adjectival form of '*P*' in this manner, then '*P*' is an adiathetic expression in the statement in question.

For example, if it is possible correctly to predicate the expression "ambitious" of Jones, so that the statement "Jones is ambitious" is true at t_1, even though at t_1 there may be no object toward which Jones's ambition is directed, then "ambitious" is a diathetic expression in that statement. Taking another example, if it is *not* possible correctly to predicate the expression "sympathetic" of Jones, so that the statement "Jones is sympathetic" is true at t_1, even though at t_1 there may be no one to whom Jones is sympathetic, then "sympathetic" is an adiathetic expression in that statement.

In the case of most psychological expressions used to give motive explanations, use of the adjectival form with the present tense reflects the classification of the expression generally, as diathetic or adiathetic. I believe "ambitious" and "sympathetic" are such expressions. However, as I suggested earlier, a few expressions may be regarded alternatively as diathetic or adiathetic, depending on how they are used. One such expression is "curiosity." Often, in explaining a person's action as something done "because he is curious," one is describing the person's present attitude—he is inquisitive about this or that thing—without thereby ascribing to him a disposition to inquire. Here "curious" is being used as an adiathetic expression; one is speaking of the person as curious only in connection with that about which he is curious at the moment. Sometimes, however, we mean to say something else. A bright lad, one who explores his surroundings and asks questions incessantly, may be said to behave in this fashion because he is curious. He has a propensity to inquire. Here "curious" is being used as a diathetic expression; one can say of the lad that he is curious without it being thereby understood that he is presently curious about something or other.

Similarly for the expression "jealousy." One may use the adjectival form of this expression to describe a present attitude—she is jealous of her sister—without thereby ascribing a disposition. But where one predicates the adjectival form "jealous" of, for example, a person who is regularly given to suspecting rivalry where none exists, one is ascribing a disposition.

Another group of expressions illustrates rather strikingly the distinguishing feature of adiathetic expressions generally—that they are always used with an object. I have in mind those adiathetic expressions in which a possible alternative to the substantival form of the expression in explanatory contexts is the transitive verb rather than the adjective. Consider, first, the contrasting case of the diathetic expression. For an explanation-sentence in which the substantival form of a diathetic expression occurs,

(1) He did *x* out of compassion.

it is always possible to set out an explanation-sentence which implies (1), and in which the adjectival form of the expression occurs. For example,

(2) He did *x* because he is compassionate.

However, some adiathetic expressions—"admiration," "hatred," "love," and "pity"—do not follow this pattern. An explanation-sentence in which the adjectival form of one of these adiathetic expressions occurs does not imply an explanation-sentence in which the substantival form of the expression occurs. The adiathetic expression "pity" is a case in point. If one is asked, "Why did she give the beggarman free meals?" one may explain why she did by using the following sentence, in which the substantival form of the expression occurs:

(3) She did *x* out of pity.

Here it is not possible to set out an explanation-sentence which both implies (3) and includes the adjectival form of the expression. The contender,

(4) She did x because she is pitiful.

quite obviously does not imply (3). Instead, the alternative explanation-sentence is one in which the transitive verb form of the expression occurs, namely,

(5) She did x because she pities him.

and this sentence implies (3).[3]

As it occurs in an explanation-sentence, the transitive verb form of an adiathetic expression highlights the distinguishing feature of adiathetic expressions, for the transitive verb must be used with an object owing to the simple dictates of grammar.

II

I have argued that the distinction between diathetic and adiathetic expressions can be described in terms of the distinct connections existing between diathetic and adiathetic properties and their objects. It may be correct to say that Jones is greedy at a particular time t_1 even though no object of a kind associated with greed is present to Jones at t_1, but it will not be correct to say that Jones is grateful at t_1 unless there are, present to Jones at t_1, objects of kinds associated (in pairs) with gratitude. Given this distinction between diathetic and adiathetic expressions, I want now to consider examples of each type as they are used to explain, first, actions in progress, and second, past actions.

Beginning with a sentence in which the adjectival form of a diathetic expression occurs, consider (6), which may be used to give an explanation of an action in progress:

(6) Jones is doing x because he is greedy.

Given (6) one can infer (7), the explanation-sentence in which the substantival form of the expression occurs:

(7) Jones is doing x out of greed.

However, from (7) one cannot infer (6). In both (6) and (7) one is asserting that Jones is acting from a motive. But in (6) one is also asserting that Jones is acting from a motive from which he is disposed to act, whereas in (7) the question of Jones's disposition is left open. Typically, one will use the adjectival form of a diathetic expression with the present tense to give a motive explanation, as in (6), only if the dispositional ascription occurring in the explicans-clause is believed to be true. If the dispositional ascription is not believed to be true, one may use the substantival form of the expression to give the explanation, as in (7). Of course one may use the substantival form of the expression to give a motive explanation even though the corresponding adjectival form with the present tense could have been used instead. For the use of the substantival form leaves open, as the use of the adjectival form with the present tense does not, the question of whether a person's behavior is being explained in terms of a disposition of that person.[4]

In the case of an adiathetic expression one is not giving an explanation in terms of a disposition in any instance. Accordingly, a sentence in which the adjectival form of an adiathetic expression occurs is equivalent to one in which the substantival form of the expression occurs. For example, (8) and (9) are equivalent:

(8) Jones is doing x because he is grateful.
(9) Jones is doing x out of gratitude.

It appears, then, that motive explanations of actions in progress reflect the distinction between diathetic and adiathetic expressions as follows. In the former case the relation between an explanation-sentence in which the adjectival form of the expression occurs and one in which the substantival form occurs is one of implication; whereas in the latter case that relation is one of equivalence.

In the more difficult case of explanations of past actions, the distinction between diathetic and adiathetic expressions is reflected in systematic differences with respect to the speaker's use of tenses. Consider (10), a sentence in which the adjectival form of a diathetic expression occurs with the present tense, and which may be used to give an explanation of a past action:

---

[3] Compare the adiathetic expression "envy," where the adjective ("envious") and the transitive verb ("to envy") both serve as alternatives to the substantival form of the expression in explanatory contexts. Here an explanation-sentence in which the adjectival form of the expression occurs and one in which the transitive verb form of the expression occurs, will each imply an explanation-sentence in which the substantival form of the expression occurs.

[4] The expression "vindictiveness" is an exception. In using the substantival form of this expression to give an explanation, one asserts that the agent is acting from a motive from which he is disposed to act. The exception comes as no surprise, since one understands by vindictiveness not only the motive of revenge but also the disposition to act from that motive.

(10) Jones did *x* because he is greedy.

On intuitive grounds one has no doubt that (10) may be used to give an explanation. That is, one has no doubt that it may be used to give an account of the problematic situation posed in the explicandum-clause (Jones's having done *x*) by way of the familiar situation presented in the explicans-clause (Jones's greed). However, one may understand that an explanation serves the purpose for which it was given, without understanding how it does. Specifically, in the present case one may ask: How is it that one can use a present tense ascription to explain a past action?

For explanations of past actions which are given by using diathetic expressions, the question can be answered along well-recognized lines. The diathetic expression, as it occurs in the explicans-clause of (10), is actually being used in a disposition-ascribing role, and the explicans-clause provides a warrant for the inference that what is true of the agent now (in the respect specified by the ascription) was also true of the agent at the time he did *x*. So, if the statement made by asserting the explicans-clause of (10) is true, one may infer that Jones was greedy at the time he did *x*. It is in terms of this inference that one can provide an answer to the question posed above. One can use a present tense ascription to explain a past action only if, from the present tense ascription, one can infer the appropriate past tense ascription.

The above analysis of explanations of past actions which are given by using diathetic expressions is avowedly Rylean,[5] and it stands in sharp contrast to an analysis appropriate to explanations of past actions which are given by using adiathetic expressions. In the latter case there is no occasion to ask how a present tense ascription can be used to explain a past action. For in an explanation of a past action given by using the adjectival form of an adiathetic expression, one will ordinarily use the adjectival form with the past tense. For example, consider (11):

(11) Jones did *x* because he was grateful.

The speaker's use of the past tense in the explicans-clause can be accounted for in terms of the nature of adiathetic properties. I have already alluded to the fact that adiathetic properties are understood in terms of the objects associated with them. One's gratitude, I suggested, is understood in terms of a pair of objects—the person to whom one is grateful and that act or forbearance for which one is grateful. When one explains an action by means of an adiathetic property, the explanation is understood in terms of the objects associated with the explaining property—specifically, the objects present to the agent at the time the agent acted. When the action in question is a past action, an explanation of it will be understood in terms of the objects present to the agent at that past time. For this reason, in giving an explanation of a past action by using the adjectival form of an adiathetic expression the speaker uses the past tense.

What of the anomalous case in which one uses an adiathetic expression with the *present* tense in giving an explanation of a past action? Suppose, for example, that (12) is true.

(12) Jones did *x* because he is grateful.

In using (12) to give an explanation the speaker is assuming that the objects present to the agent at the time the agent acted are still present to him. The case is anomalous because the speaker, in giving an explanation of a past action, is not ordinarily concerned with whether the explaining property can still be ascribed to the agent at the present time, much less with whether the particular objects associated with it at a past time are still present to the agent.

Since diathetic properties (such as greed) can be understood apart from their objects, the speaker need not use the past tense with a diathetic expression when he gives an explanation of a past action. This we have already seen in connection with (10). However, it is worth adding that the speaker may use the past tense with a diathetic expression in explaining a past action, as in (13):

(13) Jones did *x* because he was greedy.

Where the adjectival form of a diathetic expression is used with the past tense, as in (13), the resulting statement is ambiguous. On the one hand, (13) may simply be equivalent to a statement in which the substantival form of the expression occurs, as in (14):

(14) Jones did *x* out of greed.

On the other hand, it is always possible that the expression occurring in the explicans-clause of (13) is being used in a disposition-ascribing role, but with respect to an earlier span of time. If so, at the time

---

[5] See Ryle, *loc. cit.*, and also Ryle, " 'If', 'So', and 'Because'," in Max Black (ed.), *Philosophical Analysis* (Ithaca, N.Y., 1950).

Jones did *x* it would have been true to say that he is doing *x* because he *is* greedy.

It appears, then, that in giving motive explanations of past actions the situation is this: In diathetic cases the speaker has the option of using the past tense when using the adjectival form of an expression, and in that event no dispositional ascription need be understood; while in adiathetic cases the speaker ordinarily has no option but to use the past tense when using the adjectival form of an expression, for the reason that no dispositional ascription can be understood.

III

It is clear that in many cases a speaker can readily distinguish explanations which contain motives as factors from explanations which do not. For example, while it sounds odd to say that stupidity is a motive in the explanation "He did it out of stupidity," it does not sound at all odd to say that ambition is a motive in the explanation "He did it out of ambition." Taking another example, while it may be the case that Jones is tactless, and while one may have occasion to explain Jones's behavior by appealing to that trait ("Jones did it because he is tactless"), it would sound odd to say that one is thereby giving a *motive* explanation of Jones's behavior. What conceptual difference between motive explanations and other types of psychological explanation makes it possible for the speaker to discriminate between them?

Ordinarily, when speaking of a person as having done something from this or that motive, one can say much the same thing by speaking in terms of the person's wants or desires. For example, where an investment practice of buying large quantities of speculative securities can be explained in terms of greed, it should likewise be possible to explain that practice in terms of a desire to increase one's wealth. Similarly, where an act of betrayal can be explained in terms of revenge, it should likewise be possible to explain that act in terms of a desire for retribution.

I am suggesting, then, that the principal conceptual difference between motive explanations and other types of psychological explanation is that motive explanations bear a systematic connection to explanations given by appeal to the agent's wants or desires.[6] Specifically, if a motive explanation can be given of a particular action, then some want explanation can also be given of that action. Taking an example, suppose that Jones's working long hours in the office of the ABC Company can be explained in terms of ambition. That is,

(15) Jones is doing *x* because he is ambitious.

Next, consider a possible want explanation of Jones's long hours.

(16) Jones is doing *x* because he wants to be promoted to office manager.

Of course (15) may be true even though (16) is false; Jones may work out of ambition even though managing the office figures nowhere in his plans. Consider, then, another want explanation.

(17) Jones is doing *x* because he wants a good recommendation from the ABC Company.[7]

Again, (15) may be true and (17) false. However, it will not be the case that one can assert (15) and deny, in turn, every want explanation which might account for Jones's action. Speakers, in other words, will agree that the conjunction of (15) and (18) is always false.

(18) Jones has no wants which will explain his doing *x*.

Some want or wants of Jones's must explain his working long hours, lest it be concluded that he is not working out of ambition after all.

While the supposition that no want explanation can be given of a particular action rules out a motive explanation of that action, this supposition does not rule out explanations given by citing psychological factors other than motives—for example, character traits. Consider here, Jones's remarking to an officemate on how badly the latter's children are doing in school, and suppose Jones's action can be explained in terms of tactlessness. That is,

(19) Jones is doing *x* out of tactlessness.

Now read (18) in the present context. Speakers will agree that the conjunction of (18) and (19) is not always false. It may be the case that an explanation of an action can be given in terms of the agent's tactlessness, even though no explanation of the same action can be given in terms of the agent's wants. Indeed, if Jones has wants which will explain his remark to the officemate, one will not say that Jones made his remark out of tactlessness—or at any rate, not simply out of tactlessness.

---

[6] See Jeremy Bentham, *A Table of the Springs of Action* (London, 1817), for a richly detailed discussion of motives and their corresponding desires, replete with schematic tables.

[7] Complete statements of the want explanations at (16) and (17) would include clauses to the effect that Jones believes doing *x* will satisfy (or will be likely to satisfy) his want.

Psychological explanations given in terms of character traits (such as tactlessness) illustrate the class of explanations given by appeal to dispositions but not thereby to motives. Psychological explanations given by using adiathetic expressions illustrate the class of explanations given by appeal to motives but not thereby to dispositions. Indeed, the appeal to a motive will be *eo ipso* an appeal to a disposition—and likewise, the appeal to a disposition, an appeal to a motive—only in the case of an explanation given by using the adjectival form of a diathetic expression along with the present tense. Only in that case does one always assert, in giving the explanation, that the agent is acting from a motive from which he is disposed to act.[8]

Law School of Harvard University

Received March 29, 1971

---

[8] I am grateful to members of the Philosophy Department at the University of Wisconsin for a number of helpful suggestions.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 2, April 1972

IX. THE CONCEPT OF PROPERTY

FRANK SNARE

THE subject of this essay is the concept of property, not the "properties" of the metaphysicians, but the property of which John Locke speaks in chapter V of the *Second Treatise*, i.e., property in the sense of personal possessions. Indeed, had I not Locke's famous chapter, "Of Property," in mind, I would have better entitled this essay, "The Concept of Ownership" since we are concerned here with the special relationship which may hold between a person and a physical object called "owning" in virtue of which the latter may be called "personal property" or "a possession."

There has been much discussion in recent moral philosophy of "institutions" such as the institution of promise-making or of the family or of punishment, which are constituted by systems which allocate rights, obligations, duties as well as roles and offices.[1] Also much has been made of the fact that justification within such an institution or practice may be a quite different matter from the justification of the institution itself. However not much has been said, in this regard, of the particular institution of property, for this institution too allocates a complex set of rights and obligations. And this is indeed peculiar, for of all such institutions it is surely the most pervasive in our society and in our thinking. A Martian visiting our planet would understand little of what goes on in our everyday life if he missed the fact that much of our behavior is guided by, and is only intelligible within the context of, this institution. For example, he would completely miss what we are doing when we sell an automobile or give a gift or steal an apple. After all, a stolen apple doesn't *look* any different from any other apple. Of course, our problem tends to be the opposite of the Martian's. The concept of property is so fundamentally involved in our ordinary life and speech that it is difficult to step back and isolate it for evaluation. We sometimes find it difficult to raise questions as to the justification of this institution precisely because we can't quite see what it would be like not to have this institution. Locke, for example, was so incapsulated within this concept that, after assimilating all moral rights to property rights, he proceeded to read them into the structure of the universe as "Natural Law."

I

The following are examples of sentences which are on occasion used to assert property or ownership relationships:

(a) John owns an automobile.
(b) This automobile is John's property.
(c) This is his (John's) automobile.
(d) This automobile belongs to John.
(e) John has an automobile.

I say "on occasion" because some of these sentences can be used in other ways in certain contexts. For example, the personal "possessive" pronouns can sometimes be used in ways which do not imply any property relationship.[2] For example I might use phrases such as "my hand," "my person," "my labor," or "my essay" where I could have just as easily said "the hand which is attached to me and which is, in the best of circumstances, under my direct control," or "the person who is me," or "the work I did," or "the essay I wrote," where none of these imply ownership. A slave's hand is *his* hand—whose else would it be?—and yet it is his master's property and not his own. Whatever arguments there are against the institution of slavery have to be based on moral considerations and not on plays on the ambiguity of possessive pronouns. Slavery is immoral but not self-contradictory.

So I don't wish to discuss every conceivable use of the individual sentences in (a) through (e) but only that use which all of them have in common, i.e., to assert a property relationship. Notice I am assuming the reader already understands and uses sentences like (a) through (e) and words such as "property" and "ownership." Indeed, if he doesn't, he will have no idea what is the subject of analysis and he will have no basis for knowing whether what

---

[1] In particular, John Rawls, "Two Concepts of Rules," *The Philosophical Review*, vol. 64 (1955), pp. 3–32.

[2] Cf. the interesting article by J. P. Day, "Locke on Property," *The Philosophical Quarterly*, vol. 16 (1966), pp. 207–220.

I am saying is either right or wrong. If I were to tell him that "*A* owns *P*" means that *A* is to the left of *P*, he would have no idea whether that was at all plausible. A person must have an implicit (knowing how) knowledge of the meaning of such sentences if he is going to be any judge of the explicitly stated (knowing that) analyses of their meaning. This is not a paradox. Consider this analogy. One might have an implicit (knowing how) knowledge of English grammar without knowing any explicit (knowing that) rules, such as foreigners learn to learn English. As a matter of fact those books which set forth explicitly the rules of English grammar are ultimately tested by how we ordinarily speak. A grammar which had us saying things like "There is three horses" just couldn't be right. Explicit descriptions of English grammar, like analyses of the meaning of English sentences, have to be tested by our linguistic intuitions. So I do presuppose that the reader does know how to use such sentences. But I do not assume that he knows their correct explicit analysis.

We began by discussing the concept of property and now we seem to be discussing the meaning of certain sentences in English. Have we changed the subject? Not really. It is, of course, possible for a person who doesn't speak English to have the concept of property. But we shall never know whether what he is talking about is property rather than something else such as "being to the left of" until we discover whether his sentences can be adequately translated by anything like sentences (a) through (e) in English. And thus *we* inevitably end up asking, "Just what do sentences (a) through (e) mean?"

II

It is, no doubt, due to the influence of Wittgenstein that many recent philosophers have been concerned with rules. In particular, it has been noted how certain forms of behavior, e.g., making a chess move, or certain states of affairs, e.g., being in check, are possible only within the context of a rule-governed activity. I want to argue that ownership and property are possible only within the context of certain prevailing rules and conventions. Were there not these rules there would be nothing at all like owning, buying, selling, stealing, abandoning, leasing, etc. There would be persons and there would be things but there would be no property.

Now we must be careful which rules we are talking about. Some rules, although they regulate property, presuppose its existence.[3] Laws which tax property would be such as these. To change these rules would not be to do away with property in the way that doing away with the rules of chess would do away with chess. Thus we are not concerned with every rule which might have something to do with property but only with those which constitute its existence.

Furthermore another important kind of rule is relevant to deciding who is to have title to what. That is, there can be rules to answer the question "Why is *P* his rather than mine?" Different societies have had different conventions in this regard. For example, one's birth might qualify one as owner of *P*. Locke argued that mixing one's labor with *P* gave one title to *P*. Questions of distributive justice are generally attempts to justify some rule of this sort. But in spite of the importance of the issues of distributive justice, these are not the sort of rules we are concerned with. Even if some rule should tell us that it was in virtue of some characteristic *C*, e.g., mixing one's labor with *P*, that one was the owner of *P*, we might still wonder "But just what is it to own *P*?" "For *what* has *C* qualified one?" Notice that a verificationist theory of meaning is inadequate here. To ask what would show us that *A* owns *P* would perhaps involve an inquiry into deeds, bills of sale, what labor I had mixed with *P*, or whatever it is that one takes to prove *A* owns *P*. But a Martian, although he took our word that this gave *A* title to *P*, might still wonder what it was for *A* to own *P*. Notice also that, although different persons may dispute what entitles a person to own something, they are not talking about different subjects; they merely have diverse views about the same subject—property. Thus no rule of distributive justice can claim to be part of the very meaning of "property." On the other hand, it is perhaps necessary for the institution of property that there be some rules or other specifying in virtue of what *A*, rather than *B*, owns *P*; but no particular set of rules is necessary for it to be property we are talking about.

What remains then is to discuss those rules which are constitutive of the very notion of property. Our claim is that when one says that *A* owns *P* he is presupposing a set of conventions which are intended to regulate the behavior of *A*, as well as others, with respect to *P*. In a similar manner the concept of pawn presupposes a set of conventions which are intended to guide our actions in the chess game. To fail to see which moves a pawn may and may not make would be to fail to understand what pawns are. I am not saying that anyone who uses "*A* owns *P*" must actually accept the demands these conventions place upon him. There are thieves and even critics of the institution of property just as there are those who cheat at chess or else who just consider the game a waste of time. The point is that if one is going to *understand* what property is, one must see that these are the conventions involved and that if these conventions did not prevail, we would not in fact have property.

---

[3] Cf. John Searle's distinction between regulative and constitutive rules in "What is a Speech Act?" in Max Black (ed.), *Philosophy in America* (London, 1965).

I want to begin by discussing a fairly simple case of ownership: the fact that John owns an automobile. Of course a lot may depend on our choice of example if ownership turns out to be a different thing in different cases. But I choose this example because I think it is illustrative of what I consider the basic or core concept of ownership in that I suspect the various other uses of "own" will turn out to be either variations on this core concept or else understandable in the light of the analysis of this core concept. Now I want to say that it is impossible to understand this simple case of ownership unless we see, as a background, certain rules which permit certain moves and prohibit others. To start with, when one says that *A* owns *P*, for example that John owns an automobile, we must see that in virtue of ownership *A* is represented as having the following right, which I shall call the *Right of Use:*

1. *A* has a right to use *P*, i.e.,
    (a) It is not wrong for *A* to use *P*, and
    (b) It is wrong for others to interfere with *A*'s using *P*.

Thus John's using his automobile is a permitted move in the property game and our interfering with that use is a prohibited move. Now how can we test whether Rule 1 is indeed, in part, constitutive of what we mean by "*A* owns *P*"? Well, we might consider the following linguistic test. First, for (a), we might ask whether a person could assert "*A* owns *P*" and then intelligibly go on to ask whether it was wrong for *A* to use *P*. If the above analysis were correct, this would be no open question. The answer could not intelligibly be "Yes." Unfortunately, however, the above question is open and quite often in fact people do give the answer "Yes, it is wrong for *A* to use *P* even if *A* does own *P*." For example, one could quite sensibly claim that it's not right for John to use his automobile for running down people. But this gives us some idea as to how we should revise (a). Notice that what is represented as wrong with running down people is not that one is using one's property to do so but rather that one is killing people. It is wrong for a reason which has nothing to do with property. It would be just as wrong for John to use my car to run down people. So what we must say in (a) is that it is *prima facie* not wrong for *A* to use *P*, which leaves open the possibility that it may be wrong for overriding reasons. One couldn't say it was wrong for John to use his automobile precisely for the reason that he owns it. It would have to be wrong for some other reason.

We might consider this following situation as a linguistic test of (a) in revised form. Suppose John is getting into his automobile when someone challenges him with, "Who said you could use that automobile?" and John replies "Well, I happen to own it. It's mine." Now the challenger might continue in a number of ways but if he said, in a truly puzzled tone of voice, "What does that have to do with it?" we would begin to suspect some major sort of confusion on his part. He can give further reasons, either good or bad, why John should not use his automobile, but he can't, in the absence of further overriding considerations, wonder why John has a right to use what is his.

A similar sort of revision is necessary for (b). As it stands, it would suggest that we couldn't sensibly ask whether we ought to interfere with *A*'s using *P* where *A* owns *P*. But this is sometimes a sensible question. In the *Republic* Socrates raises the question of whether we ought to return weapons to a mad man even when they are his. But of course to make this plausible it is quite necessary to bring in the special consideration of the man's danger to others. That is to say, there are overriding reasons for not giving him his possessions. But one must at least go through the form of giving other, allegedly more stringent, reasons why it is not wrong to interfere with his use. Thus we should revise (b) to read that it is *prima facie* wrong to interfere with *A*'s using *P*. We might consider the following linguistic test of this revision. Suppose someone interferes with John's use of his automobile and John objects, "You can't interfere, I own this automobile." Now perhaps this person should interfere, e.g., the automobile is rigged with a bomb, or perhaps it is desperately needed to take someone to the hospital. But what is not open to the interferer is for him to reply, with honest puzzlement, "You do indeed own it, but why is that at all relevant?" Anyone who understands what property is sees why that is relevant.

However, Rule 1 is not sufficient by itself to explain what is peculiar to property rights. Ownership implies not merely that the owner has certain rights to P but that the owner is the only one who has these rights with regard to P. Property rights are in some sense exclusive. Thus we must add the *Right of Exclusion*:

2. Others may use P iff A consents, i.e.,
   (a) If A consents, it is *prima facie* not wrong for others to use P;
   (b) If A doesn't consent, it is *prima facie* wrong to use P

where there is some conventional ritual or speech act which will count as "consenting" or "giving one's permission." Notice that the addition of the "prima facie" is quite necessary. For example in (b) it might be right, in the final analysis, for us to use P even where A hasn't given us his consent. Even Locke saw that we must sometimes pull down a man's house to keep a fire from spreading or break into private granaries to feed the public in time of famine, with or without consent. The point is merely that there is a presumption in favor of not using a man's property without his consent. One must at least go through the form of presenting overriding considerations. There is something linguistically odd about the following exchange. Suppose I start to use John's automobile and he objects, "That's my automobile," and I reply by saying "Indeed it is and I'm going to use it." Now of course by Rule 1, if I thought my using it would interfere with his use, I would have to recognize the relevance of John's claim; but perhaps I know that John has no intention of using it this afternoon so that I would in no way interfere with his use. John might then say, "Regardless of whether you're interfering with my use of the automobile, I didn't give you permission to use my automobile." Now I might have a number of further replies which, justified or unjustified, would at least make linguistic sense. What I can't say is "Why is your permission even *prima facie* relevant to the issue?" unless, perhaps, I was bringing John's ownership into question. To understand what property is is to see that consent plays just this role in the property game.

It is quite possible for there to be societies where rights over objects, or certain objects at least, are not exclusive. There are tribes, for example, where many persons have interlocking rights with respect to the same object. There may be rules which eliminate conflict of use by dividing up the use in certain ways or else there may be rules which decide who shall use it in case of conflict, e.g., "First come, first served," or "Nobility first." Now of course if one is dedicated to finding "property" in every possible social system, then one might invent some entity, such as society or the community, in which ownership is supposedly vested. But this way of speaking obscures more than it clarifies. We must in any case appreciate how radically different a society is from ours where, although there are rights over objects, these rights are not exclusive. The people of such a society could not be said to own objects even though they have certain rights with respect to them.

Rule 2 also gives us some insight into the moves Locke tries to make in the *Second Treatise*. Locke's problem there concerns how it is ever possible for governmental power to be morally legitimate. Locke begins with the rights an individual has over his person, liberty and estate which, significantly enough, he assimilates to property rights. The problem then is, given these individual rights, how is it possible for the state ever to have legitimate authority over persons and their property. How can government legitimately tax property or punish criminals, for example? Locke's answer is that the only legitimate authority of government is based on consent. Locke obviously has Rule 2, not to mention certain rules we shall discuss later, in the back of his mind. But against Locke it should be noted that we needn't assimilate all rights to property rights nor must we suppose that property rights are so absolute that they can only be abridged by consent. We might consider property rights as merely *prima facie* rights which can be overriden by other moral concerns, e.g., the public welfare.

A third characteristic of property is that it may, typically, be transferred. One may sell, buy, give, and receive property. Thus we might add the *Right of Transfer*:

3. A may permanently transfer the rights in Rules 1 and 2 to specific other persons by consent,

where certain rituals or speech acts count as that consent, e.g., signing a piece of paper. Notice that there are many exclusive rights which cannot be transferred and thus would not normally be called property rights. The rights that come with my teaching position are exclusive rights I have but I couldn't sell my position or even give it away to someone else. Thus, although it is my teaching position, we wouldn't say that I literally own the position. Also there are, traditionally, a certain set of rights which are considered inalienable in that I may not sell, nor even give, them away. Thus I have no right to sell myself into slavery nor does my right to free speech involve the right to sell that right. Oddly enough, Locke held that one couldn't alienate his rights to life. What he should have concluded is that it is misleading to speak of one's person as literally one's property, although this is not to deny that each of us has very fundamental exclusive rights in this regard.

We might get a clearer idea of the substance of Rule 3 if we consider what it would be like for Rules 1 and 2 to hold between *A* and *P* but Rule 3 not to hold. For example a man might have rights of usage over a certain plot of land and indeed have exclusive rights over that land but it not be the case that he could ever sell or give the land away under the prevailing conventions. I understand a situation much like this was true in the Middle Ages.[4] A serf might be born to the land and have certain exclusive rights over it (as well as certain duties) but these rights would not be transferable. This is why it would be misleading to say the serf, or anyone, owned the land. It is clearest merely to describe the allocation of rights and duties with respect to the land.

Finally, there are several sorts of rules we can indicate only in a general way. Property typically involves there being specific rules which fall under the following general types:

4. Punishment Rules: If some other person *B* interferes with *A*'s use of *P* or if *B* uses *P* without *A*'s consent then *B* may be punished in certain appropriate manners.

Typically *B* can be punished only in accordance with certain procedures and by certain officials of special institutions (courts, judges). However Locke held this was a right each individual holds in the State of Nature but which he grants to government in consenting to join political society.

5. Damage Rules: If some other person *B* damages *P* without *A*'s consent, then certain appropriate damages may be required of *B*.

Locke held that the right to damages was not a right one gave up upon entering political society; however the right to try such cases in accordance with certain procedures was, according to him, transferred to special bodies.

6. Liability Rules: If *P* results in certain ways in damage to the person or property of some other person *B*, then *A* may be held responsible and damages required.[5]

Now rules of sorts 4, 5, and 6 seem to me to be somewhat more peripheral to the concept of property even if they are, nevertheless, important. I could imagine a society completely lacking any rules whatsoever of one of these sorts and yet having enough of the other rules (especially 1, 2, and 3) that I would say they had an institution of property even if a rather different one from ours. It seems to me the line between what rules constitute the notion of property and what rules typically regulate property is in some cases not that precise and I shall not in my analysis require a precision which is not in the ordinary concept. The really important thing is to begin to see the enormous social possibilities for different rules relating persons and objects.

III

It would be a mistake to suppose that a society either has the institution of property or else lacks it completely. The interesting differences are a matter of degree. In particular, societies differ greatly in how wide or narrow a range of situations they are willing to apply the concept of property to. Societies differ significantly in the class of persons they will allow as owners and the class of things they will count as possible objects of private possession. Who or what may be substituted for '*A*' in "*A* owns *P*" is limited by various conventions in different societies. It may be that a society will by the rules of its institutions eliminate a whole social class from the class of possible owners. Thus during the time of slavery in the Southern States a slave could not own land. In our society children and the mentally incompetent are excluded from certain kinds of ownership.

---

[4] Cf. C. B. Macpherson's discussion of customary or status society in *The Political Theory of Possessive Individualism* (Oxford, 1962), pp. 49ff.

[5] The author is grateful to one of the referees for the *American Philosophical Quarterly* for suggesting the need for this rule.

On the other hand, of even more interest to the political philosopher is the fact that what can be substituted for '*P*' is often limited by local convention. It is one of the essential characteristics of a capitalist society that one's laboring power be something one can sell, hence the possibility of surplus value in the Marxist argument. However, there are societies where no act of yours could be construed as having sold one's laboring power or else such a "sale" would be no more binding than the "sale" of one's children or of oneself into slavery would be in our society today. There are societies where one's labor is thought of as permanently tied to a particular plot of land in such a way that one's laboring power is not something one has a right to sell. Whether these ways of thinking of labor are primitive or enlightened will depend, no doubt, on the conclusions of one's political philosophy. However, it's worth noting that most political argument over the institution of property is not over whether we shall have any private property at all but over what sorts of things shall be possible objects of private ownership; for example, whether the means of production shall be things which can, by the prevailing conventions, be personal property.

IV

In considering rules 1 through 6 I am only attempting to clarify what I consider a certain core concept of ownership. However this analysis can, if applied indiscriminately, be misleading in several ways and not do justice to the variety of our ownership talk. First, I think it worthwhile to point out that one might in specific cases modify or exclude one or more of rules 1 through 6 without making it inappropriate to speak of property. For example, one might buy a non-transferable theater ticket and still consider it one's property even though it is not to be sold. But notice that insofar as the game being played involves fewer and fewer of these rules as a background it becomes less and less appropriate to use property talk.

Secondly, we shall have to admit that for certain substitutions for '*A*' as well as for certain substitutions for '*P*' the simple analysis so far given will be in certain ways misleading and thus we must be prepared to analyze "A owns P" in a somewhat special way to suit the occasion. For example, we speak not only of persons but of corporations and of states as possible owners of property. However, we should not suppose that "The corporation owns this computer" is in every way analogous to "John owns an automobile." It seems to me that the plain fact of the matter is that in a corporation we have a complex of interlocking rights few of which are exclusive or permanent (in that they go with offices and positions rather than particular persons). To speak of the corporation owning property is a convenient shorthand way of speaking of a complex network of individual rights over the object. Rather than supposing the corporation to be a mysterious person who owns things much as John owns his automobile, we should see that saying "The corporation owns this computer" is a convenient way of indicating a complex of interlocking rights that individual persons in their various roles in the organization have over the computer, none of which rights can be called property rights. In certain respects the individual within the corporation is like the individual within the feudal system in that both have offices or roles which determine rights and duties with respect to certain objects and where, in both cases, it is misleading to call the rights these individuals have property rights.

Just as taking the core concept of ownership too seriously may cause us to become mystified when intangible entities such as corporations are classified as owners, likewise we may become confused when it is certain intangible entities which are classified among the things owned. For example, what is it one owns in owning a share of stock?[6] Certainly it is not just to own a piece of paper. A child, who knew what it was to own an automobile, might be tempted to think that because we speak of owning stock, "stock" must be some mysterious entity which floats like a ghost over the paper. But stock is no more mysterious than the corporation is. We should see that "*A* owns stock" is a shorthand way of saying that *A* has certain limited rights with respect to a certain company, e.g., rights to an income of a certain kind and rights with respect to election of officers. But certainly he doesn't own the company. He has no right, for example, to give orders to workingmen.

The moral should be that, in certain cases where we use "*A* owns *P*," the specific rules and conventions within which the sentence has meaning will have to be clear from the special context or else hinted at by the kind of *A* or *P* being discussed. The

---

[6] This example is due to Benn and Peters, *The Principles of Political Thought* (New York, 1959), ch. 7. This chapter is a good introduction into the complex and diverse nature of property.

American Philosophical Quarterly
Volume 9, Number 2, April 1972

X. WHAT ARISTOTLE SHOULD HAVE SAID: AN EXPERIMENT IN METAPHYSICS

PETER MILLER

I

THIS paper is a philosophical experiment. It is an attempt to construct an Aristotelian metaphysical position which, though not Aristotle's own, contains as fundamental components categories developed, analyzed, and interpreted by Aristotle. The elements of the position are Aristotle's, but their synthesis diverges from his own. We may count the experiment successful if, while remaining faithful to the Aristotelian framework, the adjustment performed can resolve some of the difficulties within it. The resulting position may be vulnerable to criticism but, hopefully, only at points where the whole Aristotelian framework is vulnerable and at fewer points than Aristotle's own version of Aristotelianism. The value of such an exercise, successfully done, is twofold: it offers a perspective for interpreting the Aristotelian corpus, and it suggests a step that can be taken from an established position toward that more adequate account of our world for which we as philosophers are striving. And if you regard the experiment as unsuccessful, it might at least have the value of helping you, in exposing its failures, to find that more adequate position which I have sought.

A more monolithic philosophy than Aristotle's would not lend itself so readily to a treatment such as we are undertaking, in which the elements remain intact while the synthesis of those elements is challenged. But although he has considerable systematic and dialectical strength and acknowledges categories and principles that are binding in all enquiries and exemplified in all domains of reality, Aristotle is primarily a piecemeal philosopher, empiristic and analytic in temper. He isolates special topics which he probes one at a time, developing as he goes the concepts and distinctions needed to understand each problem or set of phenomena that he considers. We shall presuppose some familiarity with Aristotelian analyses of form and matter, substance and its predicaments, actuality and potentiality, essence and accidents. The point of departure for our experiment is a conclusion of the middle books of the *Metaphysics* regarding the integration of these concepts, which I think we may justifiably call Aristotle's "species monism."

Nature, for Aristotle, is composed of a limited number of species each of which has its own distinctive features and characteristic behavior, is self-sustaining and self-propagating, and is scientifically knowable. Father Joseph Owens and others have pointed out that in Books *Ζ*, *H*, and *θ* of the *Metaphysics*, Aristotle develops the doctrine that all the members of a single species owe their existence and their scientifically knowable characteristics and activities to a single persistently individual, intelligible, active, and fully actual being which is the primary substance of them all embodied in them all.[1] Rabbits, for example, are a multitude of composite substances when their form is considered in conjunction with its material embodiments. But these composite substances are secondary to the single essence they share, which as the ultimate rabbit substance is the form, actuality, and cause of them all. Aristotle is not a monist in the sense of viewing the many things of our world as expressions of a single ultimate substance. He does, however, view the many members of a *species* as expressions of a single ultimate substantial form for that species, which justifies the characterization of his viewpoint as a "species monism." A striking illustration of this is his description in *De Anima* III, 5 of the eternal active intellect, "without (which) nothing thinks," as the actuality of the soul for mankind.

The position may be contrasted, for example, with one of the views found in Plato's myths which holds that each individual living thing has its own immortal soul that is a first cause of motion governing the bodies of its successive incarnations and exhibiting a changeable character and a self-determined destiny. There may be a single platonic form for each type of thing, but not a single active principle or soul, according to Plato. A soul belongs to each individual living thing separately.[2] Aristotle, on the other hand, has wedded Plato's souls and forms to produce the notion of a single entity, at once embodied, active, and formal, which grounds the character of both the individual and its species.

---

[1] See Joseph Owens, *The Doctrine of Being in the Aristotelian "Metaphysics,"* Pontifical Institute of Mediaeval Studies (Toronto, 1963); and Ellen Stone Haring, "Substantial Form in Aristotle's *Metaphysics Ζ*" in *The Review of Metaphysics*, vol. 10 (1956–57), pp. 308–332; 482–501; 698–713.

The reasons for the Aristotelian synthesis of individual substance, form or essence, and actuality are fairly evident. He was attempting to conceive of that primary kind of being which could be relatively self-sufficient, unitary, scientifically fundamental, and complete. If it were not self-sufficient, it would be dependent on something else; if not unitary, a compound of things more basic; if not scientifically fundamental, then known in terms of something more intelligible; and if not complete, only a part or aspect of a more complete reality. But in any of these cases it would then be secondary to something else which, as its necessary condition, gave rise to it, completed it, composed it, or explained it and was therefore primary with respect to it.

Hence if there is to be a type of reality which is genuinely primary in every way, it cannot be lacking in substantial individual unity, formal intelligibility, or actual completeness. It must be the three at once, indifferently describable as a substantial formal actuality, an actual substantial form, or a formal actual substance.

Aristotle's reasoning is persuasive. The existence of anything whatsoever would seem to require the existence of beings primary in the senses indicated. Yet his synthesis of substance, form, and act is strained; their monistic unity cannot withstand critical examination; there are cracks within the concept of a single primary substance for each species. In describing how Aristotle constructed his ingenious synthesis, we shall note a few of these difficulties before proposing an alternative synthesis.

II

First let us consider the merging of form and substance. A form is the essence of all individuals of a kind; a substance is individual, a "this." They would seem to be irreconcilably different as an indeterminate universal character and its determinate particular instance. The manifest difficulties of their identification immediately embroil us in controversy over interpretations. Is Aristotle's immaterial substantial form a universal essence common to all the members of a species, a *specific form*, or is it unique and distinct for each sensible substance and thus closer in this respect to Plato's souls, a *particular form*?

My conviction is that although Aristotle operates with both meanings of "form," often equivocally, specific forms are clearly primary.[3] "Form" and "species" are, of course, identically "eidos" in the Greek text. But more importantly, Aristotle's views of knowledge and causation compel him to adopt the primacy of specific rather than particular form. For Aristotle, scientific knowledge of particular individuals in their particularity is not possible. Science cannot comprehend infinites or accidents, but deals with what is always or for the most part the case (*Met. a*, ch. 2; *E* 1027a20–26; and *Z* 1039b26–40a5). A limited number of intelligible and invariant forms is the necessary presupposition of a science of nature. Furthermore these forms are the ultimate immanent and uncreated causes of the existents and processes of nature. Change is the infectious spreading of a form from an actual embodiment to a potential receiver of it, thereby making the patient receiver become formally like the agent causing the change. The knowledgeable teacher makes the student become also knowledgeable, the hot sun heats the earth, and the parent produces an offspring of like kind. In a natural process, neither the form nor the matter is generated; only their composition in a new entity comes to be (*Met. Z*, 7–8; *Physics III*, 1–3).

The primacy of common specific forms is a central, systematically necessitated doctrine of Aristotle's. Yet *Metaphysics Z* makes clear that not just any and every common character is substantial, as the platonist is inclined to think. To bring form and substance together, Aristotle argues in Ζ that the form of an ultimate species is an ultimate, determinate unity and subject. Unlike a genus such as "animal" or an accidental quality like "white," a specific form, e.g., "rabbit," has nothing more unitary and determinate upon which it is dependent (Ζ 12). And when we look at a composite, sensible substance such as a particular rabbit or man, Aristotle notes that what is primary in it, even if non-universal, cannot be a material particular either (Ζ 3, 17). The matter of my body comes and goes as I eat, assimilate, decompose, and eliminate. Butcher it, and my body is no longer me nor human, but a heap of parts. An immaterial coordinating principle, my essence or soul, must be my primary substance, the root of my individuality and humanity (*De An.* II, 1). Might not the immaterial individual soul-substance, which is the actuality of a body, and the fundamental form defining a species be one and the same substantial reality?

---

[2] Although these characteristics of souls are common to the myths of the *Phaedo, Meno, Republic, Phaedrus*, and *Timaeus*, Plato's diffidence toward mythically expressed truths leaves room for debate as to whether he would affirm just this account taken literally. Possibly oversimplifying, then, by short-circuiting that debate, we shall take the above characterization as "the platonic view" of the soul and forms for the sake of contrast with Aristotle.

[3] Evidence and arguments for a doctrine of particular forms stemming from the ultimacy of individual substances are offered and weighed in Rogers Albritton, "Forms of Particular Substances in Aristotle's *Metaphysics*," and Wilfrid Sellars, "Substance and Form in Aristotle," both in *The Journal of Philosophy*, vol. 54 (1957), pp. 688ff.

I think not. Although the gap between substance and form can be narrowed, it can never be closed. The form defining a species must be *essentially* indeterminate; neither male nor female, old nor young, here nor there, now nor then; if it is to characterize its differing members. Yet individuals must be determinate in these ways. Nor are these individual determinations simply accidental. The controlling principle or soul of an individual controls its own body here and now and not some other of the same species there and then. Its spatio-temporal orientation, learning, development, and character make the nature of its control to differ from and often oppose other individuals which are specifically alike. For example, as Aristotle notes, the determination of moral character in men is the responsibility of the individual within that indeterminate range left open by specific human nature (*Nicomachean Ethics* III, 5). The substantial core of an individual and a common form can never merge without remainder.

III

Perhaps the union of substance and act should be more successful. Both are more determinate than a specific form could be, conforming to the "accidents" of particular matter and circumstances. Furthermore, acts and substances require each other in order to be at all. The dependence of an act upon substances is most obvious. "An act is not cut adrift from a subject, but is of *A* on *B*." (*Phys.* 202b7–8). Like the other things in Aristotle's universe, acts are not insubstantial somethings existing by themselves, but are both expressions of and resident in substances. The converse case, that no substance exists apart from act or actuality, is likewise a defensible part of Aristotle's teaching. A substance which was neither an agent nor a patient manifesting itself in some form of activity would be, in Whitehead's terms, a "vacuous actuality," difficult or impossible to distinguish from complete non-entity. An agent, of course, must always be actual in order to produce a change (*Met.* 1049b23–25). But a patient too always has some definite actual character (*Met.* H 4; θ 7). Only prime matter would be purely potential, in no way actual, but there is no such thing existent by itself in reality.

Here we face a difficulty. Normally we think of an act as doing something, and by "doing something" we mean effecting a change or making something to be different than it was before. Yet is it not the case (and surely, in Aristotle's view of the world, it is the case) that many objects are both themselves at rest and in no apparent way causing changes in others? How could such substances be either the sources or recipients of acts? Somehow the notion of act or actuality must be freed from the concept of change.

Aristotle's answer to this challenge is twofold. Although the concepts of potentiality and actuality were developed to describe the operation of the efficient and material causes, the agent and the patient, in motion and change, their use can be extended to other contexts. One way in which the extension can be effected is to note that any change caused by an agent in a patient aims at or culminates in a particular outcome or end. This is the complete actuality which the patient has the potentiality to become. The house, for example, is the complete actuality which the bricks have the potentiality to become. The process of changing a heap of bricks into a house is only an incomplete actuality. Thus it is not the *process* of change but the *outcome* which best expresses the nature of an act, and this outcome is normally in a state of rest. If we identify an act with its outcome, then the act remains when the change is left behind (*Phys.* 201a10–202a1; *Met.* θ, 1050a3–23).

Secondly, Aristotle urges a re-examination of the agent's activity itself, which will reveal acts of self-continuance as well as acts producing a change of state. Seeing, understanding, and thinking are activities in the complete sense, because they do not aim at a terminal state other than themselves; whereas reducing and learning do aim at achieving another state, thereby revealing their incomplete character. Only the former are true actualities (entelecheiai); the latter are but movements, not containing their ends within themselves. Thus an act's "doing something" need not mean producing a change. It may mean achieving an unchanging end, and the end achieved may be simply the act's self-maintenance (*Met.* θ, 1048b18–34, 1050a24–b6).

Acts and substances are never found apart from each other. Furthermore, an act may be conceived in terms of its unchanging end, possibly its own self-maintenance, which gives it the persistence characteristic of substances. Might there not be a fundamental reality at the heart of a natural being which is a substance that is an act, an act that is a substance?

Once again I think we must answer no. A substance, even when active, is more than its activity, since it is a reservoir of potential activity not yet actual. In the language of *De Anima*, it is a first actuality, an incomplete mode to be completed in and known in terms of its fulfillment as a second actuality (412a17ff.). And an activity, or actuality, even though the expression of an active substance, is more than the bare substance itself, since it is a not-always-realized perfecting state of the substance (*De An.* II, 5). Moreover, the actuality of a substance often realizes itself in different substances than its origin when a substance actualizes the perceiving and knowing apparatus of another individual (*De An.* 425b26ff.) or reproduces another member of the same species. A substance and its acts are each more than and less than the other and thus irreducibly different.

IV

The third aspect of Aristotle's synthesis, the union of form and act, would seem to be the least promising of all. A form is unchanging, eternal, never created or destroyed; whereas activity is the essence of change and motion. Nevertheless we should note the causal efficacy in time of forms and the formal character of activity. Individuals and species of a formally determinate nature continue in existence, propagate themselves, and otherwise leave the impress of their form in the natural world (*Phys.* 202a9–12). And there is no such thing as sheer, formless activity. Activities are either transitions to or continuations of a determinate, formally defined state (*Met.* θ 1050a3–b6).

But even if one should grant that there are no formless activities nor activitiless forms, the two cannot be identified. An activity always takes time and may fail to realize completely the formal end in terms of which it is identified. Even those activities which are continuations of their own state, such as thinking, are transitions to that state as existing in different, successive moments. And in Aristotle's own account, thought, that paradigm of changeless activity or actuality, continues to be only insofar as it maintains a movement from one conclusion to another. Aristotle's arguments against conceiving of thinking as a perpetual circular motion are precisely those which stress its character as a self-expending process that never repeats itself (*De An.* 407a3–35). Far from seeking its own perpetuation, thought seeks to terminate itself, to conclude, either in a practical end or in a theoretical proposition. It can continue only by adding fresh terms and making new inferences. On Aristotle's own characterization, it would seem that insofar as thinking is an activity, it does not have itself as its own end, and insofar as it is identical with the definitions and conclusions which are its starting and ending points, it is not an activity.

The tensions in Aristotle's synthesis cannot be resolved. A substance that was identical with a specific form would no longer admit of change nor of otherness to other individuals of the same species. And a specific form that was identical with a substance would no longer be essentially common and universal as required for scientific knowledge. An act identical with a form would lack the characteristic transience of activity, and a form identical with activity would lose its neutrality with respect to different spatio-temporal manifestations that further determine it. A substance identical with complete act would lose the persistent potentiality characteristic of it even when it is not acting, and an act identical with substance is, when the substance is not acting, paradoxically an inactive (first) act that becomes active in second act without any change taking place. But if substance, form, and act lose their characteristic features in the synthesis, then the synthesis, which was performed to unite their perfections in the concept of a single primary mode of being, is self-defeating.

The most that Aristotle has demonstrated in his attempted synthesis of substance, form, and act, is that each of these intimately requires the others in that it cannot exist without them. He has not shown that there are primary realities in nature which are all three identically, but only that there are none which are independent of the three. In the face of this failure, several philosophical alternatives present themselves. One is the supernaturalism of the Thomists. Noting that the formal essence and act of existence of a finite individual are irreducibly different, they conclude that only God is a complete and integrated being, an individual whose act of existence is identical with his essence and who is the cause of these being together in nature. A second alternative is the anti-metaphysical position of those empiricists who reject the explanation of any phenomena in terms of realities more primary, whether natural or supernatural. The difficulties in Aristotle's synthesis need drive us to neither of these alternatives, however. I propose a third which remains with Aristotle in developing a naturalistic metaphysics that analyzes natural phenomena in terms of primary realities encountered in nature.

V

Aristotle was led astray, I maintain, by his assumption that all the primaries of his analyses must converge as a *single* kind of primary beings. If we abandon that assumption, unsupported by his analyses, we can draw the following, more consistent metaphysical conclusions. What Aristotle should have said is: 1. Each of the three categories —substance, act, and form—is fundamental and none absolutely primary. 2. Yet each of the three stands for a basic aspect of being which is unassimilable to the others without remainder, and thus each is primary in its own way with respect to the others. 3. Nevertheless each bears intimate relations to the other two such that nothing could be at all in the primary sense of any one of the categories without also being in a secondary sense of the other categories. We thus revert from species monism to a modal philosophy like Plato's, which distinguished forms, souls and gods, and the material flux as irreducibly different kinds or modes of reality. Aristotle's distinctive contribution, however, is his exploration of the close integration of these distinct modes, which brings him nearer to the 20th century modal philosophies of Whitehead and Weiss.

This revised version of Aristotelianism can be expressed as shown in Table I using distinctions drawn by Aristotle himself but in a different systematic pattern of primary and secondary meanings. The three columns list the equivocal meanings of "substance," "form," and "act" expressing in turn the being of primary substance in the first row, the being of primary form in the second, and the being of primary act in the third. The diagonal from upper left to lower right gives the primary meaning of the terms; the other two meanings in

TABLE I

Primary senses of terms		*Senses of "Substance"*		*Senses of "Form"*		*Senses of "Act"*
Substantial Being. e.g. a man (individuates)	=	Substance as an ultimate individual subject	=	Form as instantiated: a local, intelligible "this"	=	1st Act (b):[4] individual capacities
Formal Being. e.g. "man" (characterizes)	=	Substance as formal cause, essence, or nature of individual	=	Form as species: separable in notion and common	=	1st Act (a): specific capacities
Active Being. e.g. human behavior (effectuates)	=	Substance as actuality (of body)	=	Form as actuality (of body)	=	Act as activity: exercise of capacities

---

[4] Two senses of first act are indicated. These are to be found in *De Anima*. For the general meaning of first act as the actuality of a body with definite capacities that are not necessarily being exercised cf. 412a22–28. The different senses of first act, a and b, and also act as the active exercise of capacities, are described at 417a21–b2.

each column are secondary. The three meanings in each row, while connotatively distinct, are denotatively equivalent, each referring to the same reality in the context of a different category.

The table summarizes the character and relationships of being as substance, form, and act. The first row indicates that substances, in the primary sense of that term, are ultimate individual subjects, such as a man, which "individuate" forms and acts. A primary substance's individuation of a form as its instance and of an act as its source makes the substance to be, in a derivative, secondary sense of the terms, a "form" and a (first) "act." Although a primary substance like a man is not a universal form nor sheer activity, neither is it characterless or inert. In order to be at all, a substance must instantiate a form not wholly unique to itself and act minimally with the potentiality of more fully and extensively developing and expressing itself through further activity. A man is also an instance of "man" (or humanity) and both active and able to act and develop in his own way.

Similarly the second row shows that a form is primarily the character common to a kind or species which characterizes individual substances of that species and what they can do. A form is not a platonic separate existent nor causally inefficacious, however, because it can subsist only in individual substances (as their essential nature) and their acts (as a final cause or goal) and has no substantiality and activity apart from concrete individuals and their acts. "Man" (or humanity) is the essential nature men share which their activities seek to realize. As an essence of men and a goal of their activities, "man" is also a "substance" and an "act" in secondary senses of those terms.

And thirdly, active being is primarily activity which makes substances and forms effectual as sources and goals. Activity is neither dissipated nor unintelligible because it expresses, maintains, and produces substances and instances of forms and has no existence apart from such expression, maintenance, and production. Human behavior is performed by men and realizes their characteristic human nature. As the expression of individual substances which realizes their characteristic natures, activity is also "substance" and "form" in their derivative meanings.

The three modes of substance, form, and act both stand in relation and do things to one another. While and as a substance individuates a form as its instance, the form characterizes the substance as its essential nature; while and as a substance originates its activity, the activity expresses the substance and makes it effectual; while and as an activity realizes a form, the form serves as the goal of the activity. These relations between the primary kinds of being are the inseparable converses of one another.

We have not space to elaborate how, on this account, the divine would be the coalescence or mutual assimilation of the three positive characterizations of being into a single univocal category such that any being designated by the category would be complete and simple as a substance, form, and act together. Such a univocal category would be inapplicable in fact to composite and temporal beings but applicable as an ideal that is participated in by composite beings to the extent that they are self-sufficient, intelligible, and completely actualized. Matter would be that element in the composition of beings which held them back from the ideal by making them composite and thereby dislocating substance, form, and act from one another. Of course there are difficulties on this interpretation of matter and the divine as to how the divine could be a simple unity of substance, act, and form even if there were no matter to interfere, but these difficulties are inherent in Aristotle's system itself. It is not evident that the present account, in eliminating old difficulties, adds any new difficulties, and thus it is what Aristotle should have said.[5]

The University of Winnipeg

Received June 9, 1971

---

[5] An earlier version of this paper was read at the meeting of the Canadian Philosophical Association in June 1970.

BOOKS RECEIVED

Anton, John P. with Kustas, George L. (eds.), *Essays in Ancient Greek Philosophy* (State University of New York Press: New York, 1971), 650 pages.

Bahm, Archie J., *The Heart of Confucius* (Perennial Library/Harper & Row, Publishers: New York, 1971), 159 pages.

Bartley, W. W. III, *Morality and Religion* (Macmillan and Co.: London, 1971), 72 pages.

Hall, Robert T., *The Morality of Civil Disobedience* Harper Torchbooks/Harper & Row, Publishers: New York, 1971), 162 pages.

Jaspers, Karl, *Philosophy of Existence* (University of Pennsylvania Press: Philadelphia, 1971), 99 pages.

Jones, O. R. (ed.), *The Private Language Argument* (St. Martin's Press: New York, 1971), 284 pages.

Krivine, Jean-Louis, *Introduction to Axiomatic Set Theory* (D. Reidel Publishing Co.: Holland, 1971), 97 pages.

Margolius, Hans, *Values of Life* (Pandanus Press: Florida, 1971), 117 pages.

McCloskey, H. J., *John Stuart Mill: A Critical Study* (St. Martin's Press: New York, 1971), 186 pages.

Mortimore, G. W., *Weakness of Will* (St. Martin's Press: New York, 1971), 253 pages.

Rebstock, Hans-Otto, *Hegels Auffassung des Mythos in Seinen Frühschriften* (Verlag Karl Alber Freiburg: München, 1971), 280 pages.

Rosenberg, Jay F. and Travis, Charles (eds.), *Readings in the Philosophy of Language* (Prentice-Hall, Inc.: Englewood Cliffs, 1971), 645 pages.

Rosenthal, David M., *Materialism and the Mind-Body Problem* (Prentice-Hall, Inc.: Englewood Cliffs, 1971), 242 pages.

Schiller, F., *Educazione Estetica* (Armando Armando Editore: Roma, 1971), 310 pages.

Sneed, Joseph D., *The Logical Structure of Mathematical Physics* (D. Reidel Publishing Co.: Holland, 1971), 311 pages.

Stancovici, Virgil, *Ideea de numar* (Editura Stiintifica: Bucharest, 1971), 150 pages.

Steinkraus, Warren E., *New Studies in Hegel's Philosophy* (Holt, Rinehart, and Winston: New York, 1971), 279 pages.

Stumpf, Samuel Enoch, *Philosophical Problems* (McGraw-Hill, Inc.: New York, 1971), 368 pages.

van Fraassen, Bas C., *Formal Semantics and Logic* (Macmillan and Co.: New York, 1971), 225 pages.

Vecchiotti, Icilio, *La filosofia di Tertulliano* (Argalia editore urbino: Urbino, 1971), 545 pages.

Warren, W. Preston, *Principles of Emergent Realism* (Warren H. Green, Inc.: St. Louis, Missouri, 1971), 352 pages.

White, Alan R., *Truth* (Doubleday & Company, Inc.: New York, 1971), 150 pages.

Wolff, Robert Paul, *Philosophy* (Prentice-Hall, Inc.: Englewood Cliffs, 1971), 613 pages.

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor-in-chief: JAAKKO HINTIKKA
University of Helsinki and Stanford University
Review Editor: Dagfinn Føllesdal, *University of Oslo*
Associate Editor: Risto Hilpinen, *University of Jyväskylä*
Editorial Assistant: Juha Manninen, *Helsinki*

Contents of Volume 23, No. 2, 3

Roger Rosenkrantz Inductivism and Probabilism.
Keith Lehrer Induction and Conceptual Change.
Herman Vetter Inductivism and Falsificationism Reconcilable.
Lewis G. Creary Empiricism and Rationality.
Risto Hilpinen Relational Hypotheses and Inductive Inference.
Nathan Stemmer Three Problems in Induction.
Henry E. Kyburg, Jr. Epistemological Probability.

Reviews:
M. Swain (ed.) Induction, Acceptance, and Rational Belief (Howard Smokler).
Nicholas Rescher (ed.) Essays in Honor of Carl G. Hempel (Howard Smokler).
K. Lambert and G. G. Brittan, Jr. An Introduction to the Philosophy of Science (R. Harré).
P. V. Tavanec (ed.) Problems of the Logic of Scientific Knowledge (Risto Hilpinen).

Editorial Note:
Arne Naess: Selected List of his Philosophical Writings in the English and German Languages. 1936–1970.
Books Received.

Subscription price per volume of four issues Dfl. 90 (US $31.50) including postage.

Six issues are published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY
P.O. BOX 17 DORDRECHT — HOLLAND

DIALOGUE

Canadian Philosophical Review – Revue Canadienne de Philosophie

Editors: VENANT CAUCHY *and* MARTYN ESTALL

VOL. XI—1972—No. 1

ARTICLES

First Philosophy: Metaphysics or Epistemology? KENNETH DORTER
La Démesure de la philosophie CLAUDE LÉVESQUE
Sartre et la *Praxis* économique MAURICE LAGUEUX
Moore's Account of 'Right' TOM REGAN
Why Determinism Cannot Be True CHARLES RIPLEY
Anscombe, Sensation and Intentional Objects DOUGLAS ODEGARD
Lire et délire ROLAND HOUDE

NOTES — DISCUSSIONS — REVIEWS

Subscriptions: $10.00 *a year to individuals;* $12.00 *to Libraries. Payable to the Canadian Philosophical Association in care of Norman J. Brown, Department of Philosophy, Queen's University, Kingston, Ontario, Canada.*

SABBATICAL IN OXFORD?

A comfortable, fully furnished and equipped three bedroom house located in a very pleasant and convenient residential section of North Oxford will be available for rental from mid-September 1972 to mid-April 1973. The house, which has a pleasant garden, is centrally heated and fitted with many appliances, including clotheswasher and dryer.

For details write to:

Oxford House, c/o American Philosophical Quarterly,
Department of Philosophy, University of Pittsburgh,
Pittsburgh, Pennsylvania 15213

PRINTED IN ENGLAND
by C. Tinling & Co. Ltd., Liverpool, London and Prescot

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, $6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, $6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, $6.00.

No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, $6.00.

No. 5. DESCARTES' THEORY OF NATURE, by James Collins. 1971, $6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Robert Butts
Roderick M. Chisholm
L. Jonathan Cohen
James Collins

Arthur C. Danto
James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
R. M. Hare
Rom Harré
Carl G. Hempel
Raymond Klibansky

Hugues Leblanc
Ernan McMullin
Gerald J. Massey
Richard Rorty
George A. Schrader
Wilfrid Sellars
Alexander Sesonske
Richard Taylor
Manley H. Thompson, Jr.
John W. Yolton

VOLUME 9/NUMBER 3 JULY, 1972

CONTENTS

I. JOHN LESLIE: *Ethically Required Existence* 215

II. KENTON F. MACHINA: *Vague Predicates* 225

III. BRUCE AUNE: *On an Analytic-Synthetic Distinction* 235

IV. PATRICK MAYNARD: *Depiction, Vision and Convention* 243

V. DIOGENES ALLEN: *Deliberation and the Regularity of Behavior* . . . 251

VI. JOEL J. KUPPERMAN: *Aesthetic Value* 259

VII. ROBERT CUMMINS AND DALE GOTTLIEB: *On an Argument for Truth-Functionality* . . . 265

VIII. HUGUES LEBLANC: *Wittgenstein and the Truth-Functionality Thesis* . 271

Books Received 275

PUBLISHED IN OXFORD BY BASIL BLACKWELL

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost $14.00, increasing by $1.40 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost $11.20, increasing by $1.40 for each additional 4 pages. Covers will be provided for offprints at a cost of $6 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

American Philosophical Quarterly
Volume 9, Number 3, July 1972

I. ETHICALLY REQUIRED EXISTENCE

JOHN LESLIE

I. Self-justification

Modern ethics is blocked by a dilemma, considered all too briefly in another article in the October 1970 issue of this journal. That other article—on a theory whose complexity is unfortunately as great as its importance to philosophy, the theory that the world exists because it should—asked, among other things, *what kinds of reality* ethical grounds for existence would have to be, if they were to be candidates for the task of creating a universe ("God" being the name for them suggested by some modern theologians). The dilemma concerns whether ethical judgments map any realities whatever.

I say "map," not "describe," solely because "describe" has come to be ambiguous. Example: In the teeth of A. J. Ayer's *Philosophical Essays*, such ideas as that judgments of goodness merely express emotions are held compatible with admitting that these judgments "describe facts," just because "This should be described as good," or "It's a fact that it's good," must be admitted as conventional forms of speech. Again: R. M. Hare's *Freedom and Reason*, though classifying such judgments as means of prescribing actions, claims an element of "descriptive meaning" for them, Hare's ingenious excuse being that "good," like the descriptive word "red," must be used consistently: when various features of a thing inspire you to prescribe it, you must likewise prescribe other things having those features. . . . Yet these two writers would insist that calling something good *does not map any fact or reality.*

The dilemma is this. (I) Suppose we picture ethical judgments as *mapping the constitutions of things* (where a thing = a state of mind, an act, or anything else which might be good or bad). We face the objection that no such map itself answers *whether things having those constitutions ought to exist.* This objection became troublesome in connection with whether being good was a "natural," readily detectable quality, such as pleasurableness. At first, however, its form was only obscurely sensed: people supposed that an easy refuge lay in making goodness a non-natural quality. Today, many would protest that no quality, natural or non-natural, is itself an ethical claim to existence, which is what goodness is.

(II) Alternatively, the job of ethical judgments may be of giving vent to emotions or prescribing behavior: not of mapping any reality. But now the correctness of any such judgments becomes hard to defend. We can still *affirm* correctness, classifying such affirmation as itself emotional outburst or prescription; but *defense* is another matter. We support one judgment by pointing to its logical ties with another: *if* vegetarianism is right, cannibalism is wrong: but the *if*'s plague us. The problem is not lack of knowledge, for we cannot even retain belief in, e.g., vegetarianism's moral excellence. *To be objects of belief, facts must be mappable facts,* like the supposed fact that God disapproves of pork. Yet in affirming that the enormity of meat-eating is "a fact" or "a reality," we now make no reference to cartographical exactitude, but only boo meat or prescribe its avoidance.

Let me defer proof that the dilemma's horns are sharp; my first concern is to escape them. In describing a thing as intrinsically good, might not we map some characteristic of its reality *without* mapping any part of its constitution? A thing good in itself—self-justifying, as listening to music may be—is one which itself *provides some ethical grounds for* itself or, as we say, for its own existence. (I mean that we are not forced to look to the thing's consequences for its justification; though when we do examine these consequences we may find them so disastrous that the thing ought to be avoided; listening to music can be evil when Rome burns; so "self-justifying" is not intended to suggest that the justification is always overwhelmingly strong.) Now, this provision of ethical grounds, although it is mappable, is not a mappable quality entering the thing's constitution: it is nothing like, say, Beethovenness, which helps make a musical experience what it is. Instead, it is an ethical way in which the thing is related to itself. (Compare how a man can be fond of himself; a number, the cube root of itself; etc. And please note that belief that some of my musical experiences are self-justifying, as a matter of stubborn, mappable fact, does not justify forcing the music which I like into your protesting ears, for musical experiences might be self-justifying only when pleasant.)

II. Goodness as Consequential

Early theorists picturing goodness as a matter of non-natural facts considered these facts not merely mappable but knowable: knowable not by the senses but "by intuition." Yet it is far from evident that in ethics, any more than in religion, *knowledge* should precede all activity. That one would have a duty to do nothing, if lacking certainty as to where one's duty lay, is a curious notion; for isn't "doing nothing" often a crime? Still, the early insistence on knowledge made belief in a non-natural quality of goodness appear inseparable from the strange theory that this quality could be known by what might be called non-natural means. The unpopularity of such a theory led to an unfortunate contempt for the many subtleties included in accounts of the non-natural, subtleties which I shall now consider.

When we approve a situation, its natural, ordinarily verifiable qualities are what we prize. We feel tempted to say that these *are* the situation's goodness: they are, at any rate, more than mere indicators of goodness, as redness is an indicator of heat in iron. We see no reason to suppose that truly moral beings would prize some quality utterly beyond them in the metaphysical shadows. Defenders of a non-natural quality tried, however, to allow for this. That goodness is like heat, but hidden from the five senses, is a poor caricature of the position held by, say, G. E. Moore. He insisted that intrinsic worth was *consequential*, a quality derived from those natural qualities which good men favor. Yet this perhaps saved the non-natural from crudity by making it unintelligible. Non-natural goodness "followed necessarily" from having certain natural qualities, but the nature of the necessity was confessedly problematic. (It would be disastrous, Moore urged, to connect the senses of "being intrinsically good" and "having natural qualities *XYZ*," so making the necessity a logical one. For the goodness of a thing with qualities *XYZ* would in that case follow at once from how we happened to use the word "good," and such a linguistically guaranteed fact would be as little qualified to evoke moral enthusiasm as the fact that bachelors lack wives.)

At times it seemed that the problem could be solved only by throwing the reality of value into doubt. In an early treatment of it, Moore suggested that, since intrinsic goodness *was not a part of* any thing, "good" could not stand for any *real* property—though he held simultaneously that it was both a property and non-fictitious! Later, realizing his contradictions here, he proposed that ordinary predicates *describe* the intrinsic natures of things in a sense in which ethical predicates never do . . . but precisely what sense was never explained. The remark by W. D. Ross that intrinsic goodness is never a constitutive attribute, one we must mention when distinguishing its possessor from other entities, only reinforces our puzzlement. (If non-constitutive, then in what way real?)

But what are these theorists saying if not that "property" and "quality" are unsatisfactory words here? May not the solution lie in talk of the *relation* of self-justification? For do not many relations have a reality which is *an irremovable aspect of* the related terms, but which is not *part of* these terms? (A term = anything standing in a relation.)

Some relations, it seems, are realities clearly additional to the related terms, which can exist when the relations are absent. For instance, if two boxes stand in the relation, *linked by a red rod*, then this relation can be destroyed while the boxes remain unaltered. (Someone can paint the rod green without affecting them.) But when two boxes are related in this way, that one *could fit inside* the other, the relation could not alter while the boxes remained just as before. Cases like this one inspire the comment that certain relations "are a mere matter of what the related terms are," and this invites the interpretation that such relations *do not have any reality unambiguously additional to the reality of their terms*. In effect, the most striking difference between relations such as being-linked-by-a-red-rod (which are able to change without change in the related terms) and relations such as being-able-to-fit-inside (which can change only when their terms do so) just is this: that relations of the first category obviously involve realities additional to the reality of their terms, whereas with those of the second category the presence of some additional reality is far from obvious. Thus, when two boxes are linked by a rod, there must exist not only the boxes but also the rod; whereas when they are, as we say, "linked" by the relation that one can fit inside the other, it is not clear that this brings in any reality truly additional to the boxes themselves.

Being-able-to-fit-inside is still something real, and is even, in a sense, something over and above the terms which it relates. Yet its reality *is fully secondary to or derived from the reality of these related terms alone.* There is a vast difference between how this relation links two boxes, and how a separately existing chain would link them. . . . The relation is secondary to the dyad of related terms somewhat as heat is secondary to red-hot iron: it is an aspect of them. But notice that, unlike heat, it is a *non-constitutive* aspect: its reality is entirely derivative.

Mention of such entirely derivative reality helps explain why certain relations *necessarily* hold between their terms, connecting them at all times. For any relation which could hold at one time, yet fail to hold at another, would not derive its reality from the related terms alone: the fact that its presence was not fully determined by their presence would show that its reality included something unambiguously additional to theirs.

If, therefore, intrinsic goodness is a relation of self-justification, and a relation of the derivative-reality class, a non-constitutive aspect of the related terms, then the supposedly consequential character of such goodness—the way in which the intrinsic goodness of a thing follows necessarily from the mere fact that the thing is what it is, although it is not a part of its being what it is—is explained. We can deny that such goodness is something to be prized quite separately from ordinarily verifiable qualities; for these are pictured, not as mere indicators of its presence, but as the stuff from which its reality is derived. Yet it remains true that maps of such goodness are not maps of these qualities. The way in which ethical judgments map realities intrinsic to things is, therefore, as Moore saw, an unusual one.

III. A Relation between Nature and Existence

Things good in themselves have ethical claims to existence, and this aids in explaining what is meant when being self-justifying is called a relation. For the relation can be redescribed as one which holds between *a thing's constitution* and *that thing's existence*: the first provides grounds or justification for the second.

But does not the alleged relation now stand revealed as bogus? Things are not combinations of Form and Being! There is no land of disembodied natures or essences, neither is there any refined existence without qualities. Coming to exist is not promotion from a shadow realm beyond space and time. What, then, could be intended by talk of constitutions and existences? . . . Well, consider how we decide whether to favor, for its own sake, a thing's continued presence. First we ask, "What is it?" We here think of it *as having certain constitutive qualities:* we notice, say, that it is a pleasant experience of music. Next we inquire, "Ought there to be such a thing in existence?" Here, though not recognizing any such entity as Pure Being, or supposing that existence is a quality on a par with being yellow, we think of the thing *as something real, and real in the way in which living beings and other such parts of the furniture of creation are real:* contrast, for example, the real possibility of my drawing tomorrow equilateral triangles which are equiangular, and the real impossibility of my drawing any which are not; the real possibility (or impossibility) of fitting such and such a box inside such and such another; and so on. The ethical relation between constitution and existence is therefore one which connects a thing (as having a given constitution) with itself (as having the kind of reality possessed by existing objects). It does not connect two separate entities—but it does join two terms to some extent separable in thought. (If a tiger, say, loses its existence, it is no longer anything; there is no longer any "it"; yet being an existent remains, in some wide sense, "a characteristic," one which by no means all realities have. Besides realities of possibility and impossibility, there are others which are not in their own right existents, but only aspects of existents or relations between existents. However, when one says that the constitution of a given thing provides justification for that thing *in its character as an existent,* what one is emphasizing is that there is justification for *that thing's being real in more than just the way in which a possibility is real*—or, to put this another way, justification for its being more than merely possible.)

The situation is odd enough to excuse describing intrinsic goodness as a "quasi-relation" only. "Property" certainly remains a possible alternative (especially since philosophers speak of "relational properties"). Again, one could classify such goodness as an "ethical status" of being self-justifying, a status which a thing has when it derives grounds for its existence from its own constitution. "Status" underlines what is intended when intrinsic goodness is called secondary in its reality; for we are not tempted to attribute separate existence to a status, even when it is viewed as a reality of a specifically ethical kind, a non-constitutive aspect of things.

IV. Ethical Necessity

A serious difficulty deserves consideration here. Investigations with our senses or laboratory instruments reveal only how things are constituted, saying nothing about ethical relations in which things so constituted stand to themselves. But how else could we come to understand such relations?

The difficulty would not be so serious if the affair were one merely of lack of knowledge. As already mentioned, such a lack, though no doubt a shame, scarcely provides excuse for lack of interest in morality; as in the case of religion, belief has claims to respectability here. Lack of unswerving confidence in any moral standards may lessen dogmatism in imposing them on others; yet we do believe that, say, feeding the starving is good; we have been trained to this belief; and for those not ashamed of their training, that is enough. But the main problem surely remains. It is that the unverifiable (non-natural) status attributed to goodness offends against the verification theory of meaning. The difficulties confronting this empiricist theory have decreased its popularity, yet there remains force in its contention that to understand a mapping word you must meet situations where the correctness of maps using that word can be known. For how else could the senses of mapping words be learned?

The answer, I think, comes from consideration of what inspires belief in mappable facts of value. Isn't it that men wish further support for their moral preferences—support beyond that of one preference (say, for preventing famine) propping up another (perhaps for birth control)? Values, if deserving respect, cannot be whims of ourselves or of our ancestors: the suitability of an entire network of moral aims is no more a mere matter of human choice than the correctness of geographical opinions. Men feel a need to give special dignity to their moral preferences by supposing them backed, even if unverifiably, by something beyond such preferences—something whose relevance does not itself depend on prior acceptance of a particular set of evaluative norms. Religious practices, through being referred for their justification to something beyond mere whims, gain the kind of importance which they have to those who follow them, and the same goes for the moral practices of most people. They consider that certain forms of behavior, such as acts which increase the sum of human pleasure, are, as a matter of "fact" in a strong sense, mappable fact, *required* or *marked out for existence.*

Marked out for existence? . . . To be sure, it can be morally essential for us to do something without our being effectively constrained to do it. Moral authority is often unauthoritative in practice: the ideas of ethical and of causally effective requirement are distinct. But these ideas still have something in common: it is not through chance that "necessary" and "required," "cause" and "reason for existence," have ethical as well as causal senses. There is, I suggest, a limited but illuminating parallel between requirements of ethics and of causation. When the course of events is not made crushingly inevitable by causal factors, when we are in some sense "free" to choose, then, though welcoming this freedom, we can still be somewhat troubled by it. Many of us are dissatisfied with defending particular acts by reference to free decisions to follow certain general policies. Suppose now (to take the simplest case) that some act is viewed as justified by its own constitution, rather than by its consequences: it is an act, say, of listening to music for one's own enjoyment, an act without ulterior point but which still manages to be good. We may picture it as designated for existence by this constitution, *somewhat as* various things are so designated by causally overwhelming forces; for we may feel the need to view ourselves not quite as slaves but at least as servants of factors which, though taking account of our likes and dislikes (since these may be of great ethical significance), nonetheless have a reality beyond the mere verifiable facts of such likes and dislikes.

So, some restricted but interesting analogy can be drawn between intrinsically worthwhile acts and, say, the acts of the victim of kleptomania, or the man "acting" only in the sense that his limbs are pushed so that they do things. Whether we wish to propose such an analogy depends on whether we are embarrassed at our freedom to choose between actions not crushingly foreordained. The believer in mappable facts of value does feel in some respect uneasy at free choice, theorizing that his will should freely serve something other than itself or the will of others. What he wants is factors which "require" or "mark out for existence" in the sense that they set some kind of limitation to the actions open to him, so that, while many seem available as logical and causal possibilities, only a few are *ethically possible.*

Any tendency to do good is not thrust on us by moral ordinances quite as a tendency to fall is thrust on us by gravity; some people are not at all influenced by what is ethically necessary; yet ethical realities inspire us to action through being pictured as directing our footsteps when physical duress or neurotic compulsion have failed to set up their coercive signposts. Some things we are, in an ethical way, forced to do, even when we are, in a causal way, not forced to do them; *and to accept that we have here two analogous, rather than merely punning, uses of "forced," is to understand what ethics is all about.* Called on to give the meaning of "ethical requirement," we can but speak of factors which, though we may act in utter disregard of them, do as a matter of mappable fact set authoritative limits to possible actions; for it would be circular definition to specify that these authoritative limits were *ethical*. The tightrope of analogy here becomes very slippery; but this will not be the first time that philosophy, in giving meaning to talk about an unverifiable reality, has resorted to limited analogies. Compare talk about God. (In this connection, remember that there is nothing *better* than value, nothing more qualified to inspire quasi-religious awe. Also that, like God, value is very much a mystery, as the flounderings of the experts prove.)

But cannot we demand clarity, eradicating potentially misleading references to ethical "compulsion," "necessity," or "requirement"? The intrinsically good is that whose constitution makes its existence appropriate or desirable; so why speak of required existence, choices in some weak sense "forced," etc.? . . . Well, "appropriate" or "desirable" are unhelpful because so obviously admissible. They function here as totally uncontroversial synonyms for "good"—and the worst explanatory model of anything is a precise duplicate of that thing.

A critic might protest that there is nothing to ethical necessity more difficult than this: that if we are going to stick to certain policies for action, then *that makes it requisite* to do *A* in situation *S*. *Given that* we shall follow the formulae called "moral," this *necessitates logically* that at any parting of the ways we shall select the moral path. But such an analysis would make moral requirements too hypothetical: they would become relative to the assumption that moral behavior would be pursued. What one wants is not such triviality and indecisiveness. "Ethically necessary or required" must have a meaning separable from any tendency toward causal effectiveness, but the separability must not involve reducing the requiredness to "what is essential *if*. . . ." Otherwise value becomes too much like, say, the quality of being hot. Certainly it is necessary, to the madman intent on frying the world, that he should manufacture hydrogen bombs; it is "a thermal requirement" that these bombs should be exploded; but the requiredness of good things—their being marked out for existence—is surely a more integral part of goodness than such a parody implies.

Lest these remarks seem too puzzling, let it be observed that those who speak of ethical *prescriptions* show themselves at ease with the idea of *requiring or marking out for existence:* their position differs from mine chiefly in this, that they see a man's acts of prescribing something as marking it out for existence, instead of picturing the thing's own constitution as performing that function. In explaining what prescriptions are, we can point out that they are satisfied when the actions which they mark out for existence actually come to exist, so that assenting to them is a matter of carrying out such actions. The ethical requirements which I envisage are, similarly, ones satisfied by the actual existence of things. Notice that this supplies a readily intelligible basis for a comparison between ethical and causal requiredness; for the satisfaction of a causal requirement likewise involves the existence of whatever is required.

V. Having a Claim to Existence

From here on I assume that mention of "required existence" is helpful in ethics. One can say, for instance, that the goodness or self-justification of any intrinsically good thing is a relation of ethical requirement connecting its constitution with its existence. Yet this invites the following awkward queries. (1) When a thing exists, where is the sense of saying that its constitution merely makes its existence required? For the thing has no mere ethical claim to be present, but actually *is* present. (2) When, on the other hand, the ethically required does not yet exist, between *what* and *what*, in this empty situation, could any relation of requirement hold?

But cannot one reply as follows? (1a) The ethically required remains the required even when it exists. Good, once achieved, fails to become insignificant; its actual existence importantly answers an ethical need, and this makes it essential to guard against its disappearance. Remember also that the constitution of an actually existing thing might make its *absence* needful, as would be so if the thing were intrinsically evil. How, then, could there be anything absurd in speaking of the ethical need for such an actually existing thing to be *present*? (2a) In the case of not-yet-existents, talk of an ethical requirement does not involve belief in a relation joining the constitution and existence of any actual thing. Instead, *the existence which a thing of a certain sort would have* must be pictured as in the relation of being made ethically required by *the constitution which such a thing would have*. This is not to say that the two terms of this relation, together with the relation itself, actually exist, for everything here is as hypothetical as in the case of the equilateral triangles which I may draw tomorrow (triangles whose equilateral character must be pictured as in a certain relation to their equiangularity, namely, the relation of requiring it logically). In particular, the constitution of a merely possible thing is no more than the being-of-a-certain-sort which *would be* the case with a thing of a certain description, *if* there were to exist such a thing. But though it is only hypothetical matters which are in question, they are of actual significance, for the absence of a good is not to be treated lightly. The way in which the existence which some describable possibility *would have* is related to the constitution which it *would have* is an affair of real ethical importance, just because the relation is one of making ethically required. And this importance is what we have in mind when we speak of real claims to existence which mere possibles possess.

These comments are made essential by professional attacks on what is described as "ought-to-existness" (see, e.g., chap. 5 of H. A. Prichard's *Moral Obligation*). It is suggested that talk of this odd attribute involves belief in a half-way house between existence and non-existence, a realm of semi-realities clamoring for full reality. But this forms no part of my position. Either a thing exists, or it does not; yet that does not make "being what ought to exist" into an impossible third alternative. When, for example, a thing possesses an ethical claim to the existence which it has, a claim which bad things do not possess, then we have here no "third alternative," but a bonus.

VI. UNCONDITIONAL CLAIMS TO EXISTENCE

Having pointed out that the need for good things to exist transcends the matter of whether there actually exist such things, let me now suggest that it transcends also all other matters of actual existence. To begin with, it can precede people able to act in support of it. This rules out a prescriptivist approach, which analyzes talk about needs as urging people to do things. Again, it conflicts with any extreme deontology: I mean, any approach analyzing the good, without residue, as that which people have a duty to do or to produce.

But does not my own strategy, of drawing an analogy between ethical and causal requirements to give meaning to the basic concepts of ethics, make duties—needs for action—fundamental? (For it is with ethically and causally compelled actions that the analogy is concerned.) . . . No, for perhaps the only way of deciding which actions are required is to consider a requiredness possessed by things other than actions. Admittedly I begin by comparing the man who cannot avoid various bodily movements with the one who sees similar movements as (obviously in not quite the same sense) forced on him by ethical factors. But may not an arm, pushing the limbs of the man under causal compulsion, itself be a limb which is pushed? And when, in causation, the requiredness of one reality can be derived from that of another, may not the same be true in ethics? Necessity spreads from one step to another in a logical proof, or from the movements of one carriage to that of another when a train begins its journey. Now, many actions can seem to owe their goodness to something similar: the ethical necessity for *things other than actions* extends to cover the actions which yield them, like dawn diffusing across a sky.

Many things which it is natural to call ethically needful or desirable could not be so according to theories which make needs for action basic to all such desirability. How would these theories analyze the assertion that it was indeed desirable that the laws of nature were such as to allow life to develop? Or, that an ethical need was answered when living beings came to possess freedom of action? Those are not affairs which we living beings can control by our free acts! But must we then conclude (as prescriptivism or any extreme deontology would, apparently, wish) that there was nothing needful here? More seriously, imagine that someone can wipe out life with nuclear weaponry, and suppose that the process would amuse him. If all ethical reality is ultimately tied to the fact that people should behave in certain ways, then we could not condemn his fun on the grounds that it would be succeeded by *an unfortunate lack of life*. For to say that a lack is unfortunate is to attach ethical importance to it. Now, the hypothesis is that "ethical importance" makes sense only while there are people alive. Hence absence of life, once produced, would have no importance: it would not even be importantly inferior to the position which it replaced since no one would be present to possess a prescribable duty to rectify it. We could not declare that, given a total absence of people, there would be an importantly unsatisfied need for a good world of people to exist. Yet surely we must be allowed to declare this.

A final step remains. Ethical needs may look so hard to kill that they would survive, not only the absence of people, but even the emptiness of the entire class of existents. Consider the case of a mechanism able to annihilate absolutely all realities normally placed in this class. After doomsday the unanswered need for a good world would surely be something real and important, though nothing would be present to record its reality and importance. The unconditional significance of possible goods is a reality without being an existent; it is a reality precisely as the *in*significance of such goods would be a reality if they were insignificant (and presumably no one would want to make *that* an existent). A blank universe could be an important tragedy, though there would be no one for whom it was important: a tragedy not, of course, in the sense of being worse than nothing, but by comparison with all the good that there might have been.

The tragedy would be, not an existent, but a tragic lack of existents.

VII. First Main Alternative: Naturalism

Moore was, I believe, right in maintaining that failure must greet any "naturalistic" attempt to define intrinsic goodness in terms of qualities verifiable by our senses or in laboratories. For these are constitutive qualities, ones which make things just what they are, whereas intrinsic goodness is a relation of self-justification, or status of having required existence, consequential to a thing's being just what it is.

There have, however, been attempts to revive naturalism on the excuse that Moore's procedure, that of taking single natural qualities and establishing that their goodness was questionable, overlooked the possibility that value involves more than just one quality or simple mixture of qualities. "Good" may connote some very varied set of conjunctions of pleasure, knowledge, courage, creativity, and so on. Demonstrations that each one of these things can have dubious goodness—the torturer's pleasure, for example, being often judged an addition to the intrinsic evil in the torture-chamber—are therefore inconclusive.

When challenged to give a precise account of the chemistry of value, the naturalist complains that he is asked to solve all moral problems. He portrays himself as making only the humble, obvious point that "good" does not stand for an invisible, non-natural quality, but is a label to whose applicability many (but by no means all) natural qualities are relevant, and guaranteed relevant *by ordinary language.* To doubt, say, that absence of war is always *prima facie* good, would be like doubting that all wives have husbands.

This account of ethical language is, I suggest, quite wrong, for there is nothing ethically conclusive in the mere fact that a word, "good," is ordinarily used of situations of specified types. *Some* facts about use of this word are indeed important: they tell us what subject ethics is. But the most important is surely this: that *good* grounds for action are conceived as *adequate* grounds. Therefore one must take the naturalist to argue that "adequate grounds for action" *has the sense of* "grounds concerned with producing some at least of the things *ordinarily called* good: namely, knowledge, blessed ignorance, joy, noble anguish, calm, excitement, self-indulgence, self-denial, . . . , ..." Which is messy.

The more he tries to reduce the mess by drawing up rules for weighting the influence of pleasure, knowledge, health, etc., on goodness, the more idiosyncratic becomes the sense he assigns to "good." Again, reference to human wants could be thought to get him nowhere, since either one defines "a want" as an ethically important lack—something whose banishment *is good*—or else one associates wants with actual desires, facing such difficulties as that one fairly common desire is to torture for fun. Admittedly, torture is not desired by its victim; yet if one defines goodness as what *most people* desire (or, perhaps, would desire if they had experience of it), then this ethical tyranny by the majority trivializes the supposed goodness of giving most people what they desire: it reduces it to tautology. In effect, the mess follows from including in the sense of "good or adequate grounds for action" various matters which ordinary users of language might very well consider to provide good or adequate grounds, but which they would not see as guaranteed to do so by the mere meanings of words. It must be at least meaningful to deny, in opposition to naturalism, that the grounds for action ordinarily called good are anything like good in fact. Only the meaningfulness of such a denial—the fact that it does not first say something and then withdraw it, as a contradiction does—can allow its opposite (the assertion that these grounds are by and large ethically satisfactory) to be a source of inspiration; for undeniable assertions, mere tautologies, are uninspiring. Any rules of conduct *whose sole claim to authority was one based on mere rules of language* would have no right to our respect.

If to deny the goodness of a generally accepted rule of conduct were *meaningless*, then it could not be *morally despicable*—which, in my belief, many such denials are.

But is there really any way in which naturalism can be refuted? I have attacked it as a quasi-sociological report on ethical language, unable to encourage good actions. (Moore's caricature of it was: "Do, pray, act so, because the word 'good' is generally used to denote actions of this nature.") The presupposition here is that ethical philosophy ought to provide inspiring reasons for acting, reasons which are, I suggest, given by the hypothesis that certain things are required, marked out for existence, as a matter of specifically ethical (non-natural, unverifiable) reality. But anyone priding himself on down-to-earth beliefs might protest that no amount of philosophical talking could be expected to make a man saintly. And evidently there is something in this: a philosopher holding a correct theory of ethics could still be wicked.

What I find disappointing in naturalism, however, is not simply its fully excusable failure to guarantee moral enthusiasm. Instead, it is that it fails even to provide anything which might reasonably evoke enthusiasm. What could be more damping than such a naturalistic thesis as that *the factor unifying the many and varied qualities which make things good is that men ordinarily call "good" those things which have these qualities?* Or, *that what makes it ethically required to give people what they want is this: that "ethically required" means "required if one is going to do good," while "doing good" is just a puzzling way of saying "giving people what they want"?* Or, *that to call misery the supreme good is only to commit an error parallel to claiming that squares have six sides—an error concerning language alone, and not the world which language describes?*

VIII. Second Main Alternative: Prescriptivism

As discussed earlier, prescriptivism denies that to call a thing intrinsically good is to map any aspect of it. Thus, unlike naturalism, it successfully recognizes the difference between attributing certain constitutive qualities to a thing and saying that the thing has an ethical claim to existence. However, it accounts for this difference, not by interpreting ethical claims to exist as realities of a non-natural, specifically ethical kind, but by classifying the statement that such-and-such has a claim to exist as an act of recommendation or prescription.

This faces the previously sketched objection confronting any theory which makes needs for action basic to all ethical desirability: the objection that no such theory makes adequate allowance for the ethical status of some affairs which we should surely want to judge ethically. It would be ridiculous to *prescribe* that we should, or should not, create good universes, or that we should ensure that the scientific laws which our universe obeys are laws which allow life to exist. But even more crucial is this: that the prescriptivist has severe problems in showing that he is not acting arbitrarily in preferring any one set of moral standards to any other. (The same goes for supporters of similar theories, from emotivism onward, whose differences are small by comparison with their agreement that ethical judgments do not map realities.) In attempting to counter the charge that he reduces goodness to arbitrariness, he may insist that a prescription has full ethical force only if it is in line with a well-integrated system of further prescriptions with which the prescriber could, if required, specify a policy for living. Stress may be laid on a need for the aims which a man prescribes for himself to be much the same as those he prescribes for others similarly situated. But must not our evaluations look for support to something quite different from the way in which they reflect a code for living, even a code whose principles prop one another up, and which is prescribed to all men? Isn't it the correctness of such a code which we wish defended—defended, if not by *proof* that its ethical appropriateness is a matter of mappable fact, then at least by the *supposition* that it is so?

The difficulty is that a prescriptivist can find no effective defense for his moral standards in the claim *that they guide toward good things*. On his analysis, such a claim can be unpacked into a set of recommendations of the things which the standards favor, recommendations saved from accusations of arbitrariness only by appeal to the standards themselves; hence, though he can of course *make* the claim, to see it as *effective* in defense of those standards invites charges of circularity. Yet isn't it essential that a complete set of standards should look for effective support in the supposition (whether provably correct or not) that what those standards favor is indeed good, i.e. worth favoring? (No doubt any prescriptivist can insist that the things prescribable in the light of his standards are absolutely worthy to be favored, worthy in a way which is not merely relative to any man's acceptance of those standards; what he here makes clear is that he is prescribing that absolutely all men should favor those things; but my point is that such insistence, coming from him, neither gives any warrant for the standards, nor suggests where one is to be found. For this whole-hearted act of prescribing things adds nothing new to the situation: it is a long-winded way of prescribing acceptance of the standards, rather than a justification for them.)

Prescriptivism's acts of ethical grading are thus very unlike, say, the farmer's acts of grading apples with the help of standards which are a guide to—and so can be justified by reference to—the price they are likely to fetch. The grades which the prescriptivist (or emotivist, etc.) gives do reflect his standards, so perhaps cannot themselves fairly be called arbitrary; but are not the standards arbitrary? (Plainly, an opponent of prescriptivism cannot prove that such and such standards should be adopted merely by asserting that the things favored by those standards have, as a matter of mappable fact, claims to existence. Yet in asserting this he does at least postulate a source of justification for the standards, i.e., an area of realities which would, if present as postulated, make the standards appropriate. Compare how maps of hell may help to encourage moral behavior although their correctness is unprovable.)

Prescriptivists can still provide a blow by blow account of moral disputes. Unable to prove that the blacks and whites of my own ethical maps are correctly positioned, I appeal to principles which the man to be persuaded already advocates; now, this the prescriptivist also can do. It can seem to him, therefore, that interpretation of such manoeuvers as cartographer's salesmanship is gratuitous: the aim of talking morals is to get people to do things, not just map things. . . . Yes, but one trusts that the talk will get them to do *the right* things. To say that something *really ought to be* prescribed—Is this merely to prescribe it on the grounds that it really does fit standards which are themselves prescribed? Would it not be more morally inspiring, would it not exert greater prescriptive pressure, if it expressed a belief in mappable facts of ethically required existence?

IX. Postscript

Perhaps we have here no utter defeat of naturalism and prescriptivism: perhaps what I am in the end saying is that the world as these doctrines conceive it includes nothing which I personally should consider adequate grounds for action. The naturalist, in saying that facts about the normal usage of "good" guarantee that a particular way of life is good, need not see himself as giving *any reason* for choosing that way of life, for if he is sophisticated he will see that mere linguistic findings are not themselves reasons for action. The most that might be thought is that they ensure that certain reasons, such as the wish to spread happiness, are with linguistic propriety named "good." It is, he may say, up to each man to have a taste for goodness, and if he does not have it we can but brainwash, jail, or shoot him. Likewise the sophisticated prescriptivist, in saying that a whole way of life is "in fact" good, may hold that he is not outlining *a reason* for following it. Instead, he is prescribing that all mankind is to follow it. Now, it seems to me that it should be embarrassing to confess, in this way, that *to call the goodness of a way of life a matter of fact is not, in itself, to supply any reason for following it* (or, which surely comes to much the same thing, that it supplies one, but only to those already intending to live in a good way, or eager to have a way of life prescribed to them). This is because I think that, on the contrary, there is something here which has considerable ability to reinforce moral enthusiasm, if any traces of that enthusiasm are present. This is, I suggest, revealed by the analysis which presents certain ways of life as required or marked out for existence as a matter of fact which (a) is mappable and (b) concerns more than the question of what standards are employed by ordinary users of language in deciding where to affix the label "good or required." . . . But unfortunately I cannot simply shout that *truly moral* enthusiasm for a way of acting finds its justification in such a matter of fact, for it is this which is in dispute.

I myself like to think that standards for ethical grading, like standards for churchgoing, could find their warrant in realities entirely beyond all such standards (even those adopted by the ordinary man); but if an opponent is not attracted by this theory, it is pointless to argue that, *as a matter of ethical fact as I conceive it*, he *ought to be* attracted. It is, certainly, tempting to urge that no one could wholeheartedly reject the specifically ethical realities which I have struggled to describe, since to do so would be to think life pointless and fall into utter depression. Yet, when one thinks it over, how incongruous such depression would be! As if the absence of any factors able to give to behavior what I should accept as *ethical appropriateness* would make it *ethically appropriate* for me to sit depressed! Men, in effect, are logically incapable of living without pursuing some policies for action, since even to throw up one's hands in despair is to act somehow. And when we have a logical necessity for adopting some course of action or other, when we are absolutely compelled to choose somehow, then (the naturalist can ask) why require us to put up a defense of our choice to act in the ways ordinarily called "good," apart from the circular defense of remarking that ordinary language guarantees that it is good to act so? Or (the prescriptivist can ask) why hesitate to prescribe acts with the help of ethical words, given that we intend to conform with ordinary language by directing the prescriptions at all men, ourselves included?

What I think wrong with such a way of thinking is that it cites too sledge-hammer-like a reason for accepting some sets of aims. Admittedly, if I were to lose faith in the specifically ethical reality which I call "being marked out for existence," I should still find myself *logically compelled* to select some way of behaving; but this means that I should be thus compelled in all logically possible situations, even ones in which all available options were in fact equally good or bad; and therefore, it seems to me, such a ground for action, though in a way impeccable, is in another way unsatisfactory. . . . Yet if you find it satisfying, I have little hope of converting you, for, as discussed, my attempt to show that you *ought to* feel dissatisfied would have to assume the correctness of linking *oughts* with realities of requiredness of a sort which naturalism and prescriptivism deny. These doctrines cannot be crushed, but only abandoned.

University of Guelph

Received July 16, 1971

American Philosophical Quarterly
Volume 9, Number 3, July 1972

II. VAGUE PREDICATES

KENTON F. MACHINA

EVER since Aristotle, most philosophical investigations into the relation between language and the rest of the world have yielded constructions in which predicates in the language function in an artificially precise fashion.

Typically, for each predicate in a given formal language it is assumed that we are also given some property (or set) or some relation (or set of ordered *n*-tuples), in such a way that the predicate can be truthfully predicated of objects if and only if those objects possess the property or relation in question.[1] It is usually also assumed that should the objects fail to possess the requisite property or relation, the relevant predication is not only not truthful, but also it is false. One need then only add that an object either possesses or does not possess any given property (and that objects either stand or do not stand in any given relation), to arrive at the conclusion that any well-formed meaningful predication in the language is either true or false, without other alternatives. In this way, a certain sort of precision is built into the operation of the language, since one can in every case answer the question whether a given predication is true or whether false with a simple "yes" or "no." It is this precision which I find seriously artificial and misleading.

I shall argue that the predicates of a natural language do not always have that sort of precision, that it is possible to construct coherent systems of formal semantics without including this kind of precision, and that this possibility tells us something about logic and about the way language can relate to the rest of the world. In particular, the classical formulation of the principle of bivalence which claims that every proposition is either true or false does not state a logical truth; it states a falsehood, and stands in need of repairs.[2]

I.

Even if every predicate in English were precise, it would not follow that no language can contain imprecise predicates. So, we could concentrate on constructing an artificial language containing at least one imprecise predicate expression; in fact, ways of doing this will be indicated in Sect. II. But English itself already contains imprecise predicates, e.g., "red."[3]

Let us consider the case of "red" in detail. Imagine a color chip of orange-red hue, emitting light of wavelength 6465Å (in the middle of the borderline region between clear reds and clear oranges) if you will, present in a normally lighted room under normal observation conditions. Let S be the sentence, "This chip is red," referring to the color chip in question. Is S true? False? These queries seem hard to answer with any certainty. Presumably it is not hard to see that cases like this one require some sort of theory to account for the difficulty people generally encounter in classifying such colors. It is my contention that a full accounting must include the fact that "red" is imprecise, and that this imprecision renders S non-bivalent. (Of course, I also think there are many other imprecise predicates in English.)

Various alternatives are open to the philosopher who wishes to preserve bivalence in the face of such cases. First, there is the *no proposition theory*, which claims that S does not express a proposition at all, and that hence S is neither true nor false. The theory I have in mind here is not concerned with denying the existence of propositions in general; it only claims that S is special for some reason, in that it looks as if it expresses a proposition when it really doesn't. (This theory could be expressed without using the term "proposition" at all, but the correct substitute term would depend on just what one takes the truth value bearers to be, if not propositions.)

---

[1] E.g., Rudolf Carnap, *Meaning and Necessity*, 2nd edition (Chicago, 1956), p. 5. In Carnap, and elsewhere, however, some attention has been given to so-called "reduction sentences" which in effect destroy any simple relationship between properties and the predicates thus "reduced." The theory of reduction sentences thus forms one exception to the typical pattern.

[2] It is possible to deny the principle of bivalence while still accepting something that looks like the law of the excluded middle. E.g., Bas Van Fraassen, "Presuppositions, Implication, and Self-Reference," *Journal of Philosophy*, vol. 65 (1968), pp. 136–152, proposes a system in which "$P \vee {\sim}P$" is a theorem even though bivalence fails. I shall confine myself to discussing bivalence. Without making any ontological commitments to the independent existence of abstract entities, we may adopt for convenience a way of speaking which appears to refer to propositions.

[3] Since English is a rather complicated, incestuous family of dialects, some of which are no doubt lacking some of the family resemblances which will come under consideration in this paper, my claims should be taken to be claims about semiformal American English.

To such a theory I can only reply that I shall *give* the "no proposition theorist" the word "proposition" to do with as he pleases; he agrees with me that S does not express something which is either simply true or simply false, and he does not manage to exempt the logician from dealing with whatever it is that S expresses here, whether under the title "proposition" or not: S still has normal implications; it can be contradicted, negated, and conjoined to propositions.

For myself, I shall continue to use the word "proposition" (without ontological commitment), and to say S (usually) expresses a proposition. S can be used to assert of our troublesome color chip that it is red — a proposition which would have been simply true or simply false, without complications, if only the chip had been of a different shade. We shouldn't have to know what shade the chip displays before we know whether S expresses a proposition. It is sufficient to know that S predicates a perfectly coherent predicate of an actually existing object denoted by the grammatical subject of S.[4] The logical coherence of the predicate term in S is demonstrated by the fact that it is possible for actually existing objects to be constituted in such a way that the predicate term of S can be predicated of them *with truth.*

So much for the "no proposition theory."

Another alternative open to the defender of bivalence and precision is the *one proposition theory.* This theory claims that there is exactly one bivalent proposition which S as a sentence of English can correctly express in any one set of circumstances which would interest us here. Our common difficulties over deciding whether S is true under such a set of circumstances must then be merely epistemological: Either we are ignorant of the facts, or we are ignorant of the proposition expressed by S, for if we knew both, then, so the theory goes, we would know the truth value of S. One can generate different varieties of the "one proposition theory" by giving differing analyses of the one proposition which, according to the theory, S correctly expresses, or (what amounts to the same thing) by giving differing accounts of the truth conditions for S. But so long as color properties are thought of as properties which a thing wears, so to speak, on its sleeve, it does not seem plausible to suppose merely that there is some color property, called redness, which the chip either has or doesn't have, the possession of which would be necessary and sufficient for the truth of S. This supposition lacks plausibility because if it were true, it is hard to see why we should have so much difficulty with our color sorting.

When one casts about for a rendering of S that will make the "one proposition theory" work, wavelength considerations suggest themselves. Perhaps, it may be thought, S says that the predominant wavelength of reflected light from the chip falls within a given range of wavelengths —namely, the red range. If S says this, then it is easy to understand why we have difficulty with some color chips but not with others: We take the color appearance of the chips as *evidence* for their wavelength emissions, using scientific reports perhaps as a guide, but in some cases, when the color appearance becomes orangeish, we are no longer sure whether the dominant wavelength falls within the proper range. But recall that the "one proposition theory" claims there is exactly one bivalent proposition correctly expressed by S in a given set of circumstances. Thus, if the appeal to wavelength is to do its job, the boundaries of the red range will have to be precisely specified.

Obviously, one could, if he were stubborn and independent enough, insist on establishing a fixed range of wavelengths as the red range. Or, one could establish anew a fixed range of wavelengths for each occasion when the truth of S was in question. That would be a way of introducing "scientific" precision into the language, and it might even catch on, at least among physicists, so long as our theories about the character of light don't change too much. Nevertheless, it is patently false that such a reading of S is the *only* correct one under such circumstances as we have imagined. The standards which specify what S may correctly express come from the community, not from the inventor of scientific precision.

There are other imaginative proposals for reading S in such a way that the question of its truth with respect to our color chip gets settled. One might suggest, for example, that the truth of *S* is determined by whether or not a majority of the adult native speakers of English would classify the chip as a red one if they were to give a considered, honest opinion. Or, perhaps instead, one might suggest that the particular reasons we may have for wanting to decide in a particular case about *S* determine the one precise proposition expressed by *S* in that case.

---

[4] Those who believe (as I do not) that there are logical types for predicates will want to add that "red" is also of the proper type here.

All of these proposals suffer from the same malady that was fatal to our first attempt to introduce precision into the language: There just is no reason to suppose that in the specified circumstances *S* can correctly express *only one* of these propositions. But the "one proposition theory" claims that there is always exactly one bivalent, precise proposition which can be correctly expressed by *S* in the circumstances. Moreover, the more bizarre the rendering of *S*, the more likely it becomes that in fact *S* in English, spoken literally, is not *ever* in fact used to express the proposition claimed for it.

At this point, it appears that the insistence of the "one proposition theories" on the existence of one and only one proposition correctly expressible by *S* makes such theories unworkable. This appearance generates a suggestion for the defense of bivalence and precision: abandon "one proposition theories" in favor of some sort of *many proposition theory* (an MPT. An MPT will claim that *S* may correctly be used to express more than one proposition in any given set of circumstances of interest to us here, and that each of these propositions is bivalent. Some of the propositions suggested by the imaginative versions of the "one proposition theory" might do. Our difficulties with the orange-red chip can then be explained: This chip falls into some, but not all, of the various extensions of "red" corresponding to the various propositions expressible by *S*. Thus, *S* is true if it is taken to express a properly chosen proposition about the chip, and false if taken to express some other suitably chosen proposition about that very same chip.

Under such circumstances, it is reasonable to suppose we would have trouble keeping the various propositions straight, and that we might be undecided about whether *S* should be taken to express one of the true propositions or one of the false ones.

Now, I of course suppose that our difficulties with the orange-red chip do stem from some sort of indefiniteness about the extension of "red," but I do not believe that an MPT can provide an adequate account of this indefiniteness.

To assess in greater detail the failings of the MPT it is first of all important to catalogue the standard jobs which propositions are supposed to do. (Similar remarks apply to nominalistic substitutes for propositions.) Propositions are the truth bearers for our utterances, in that it is only by expressing some proposition that an utterance gets its truth value. The utterance gets the truth value of the proposition thus expressed. Moreover, propositions are the relata of logical relations and the objects of logical operations such as conjunction, implication, and negation. Thus, propositions are asserted and denied by utterances made by people, and the implications of a given utterance are determined by the implications of the proposition expressed by that utterance. Moreover, it is usually thought that propositions are what we believe, know, and disbelieve. Such are the propositions of the MPT.

Given this understanding of the MPT we are in a position to see that part of it is probably true: *S* can probably be correctly used to assert more than one proposition in roughly the sort of circumstances we have in mind, without any gross violations of common English usage: e.g., presumably one might choose to say something with implications about wavelengths when one utters *S*, or one might choose not to do so. But what is *not* yet clear is whether these correctly expressible propositions are all bivalent, with a precise extension given to "red." I wish to maintain they are not all bivalent — in fact, I believe *S* does not *normally* express *any* bivalent proposition. It is this point that needs discussion.

Notice that we would have trouble classifying more than one orange-red color shade, and a decision about one would not necessarily settle the question with respect to all the others. To account for the phenomena, then, the MPT will have to maintain that there are more than two allowable extensions for "red" — in fact, we shall need quite a few such extensions — presumably at least one for each discernible color difference that falls within the problematic range. In this way, we find ourselves confronted with a sizable array of propositions each one of which could quite properly be expressed by *S*, according to the theory under discussion, some of which are true and some of which are false.

I believe we are now in a position to see the oddity of the MPT. The theory's first unhappy consequence is that hardly anyone normally believes *anything* to be red. To obtain this result, we need the factual (and certainly reasonable) premiss that normally hardly anyone has settled on a precise boundary for the extension of "red" when he does what the nonphilosopher would call believing that some given thing is red. Here it doesn't matter whether the given thing is a clear case of a red thing, or not. Now, assume both the MPT and this factual premiss; furthermore, place propositions in their traditional role as the objects of belief. It then turns out that in order to have a belief that a given object is red, one must believe some proposition which employs a precise extension for "red." But *which* proposition is that one? None will do, normally, according to our factual premiss. Hence the result that when a nonphilosopher thinks he believes the fire engine is red, he is mistaken, unless his belief includes a precise extension for "red" — which is to say, the nonphilosopher is normally mistaken about this! At least so the MPT would have it.

The second sad consequence of the MPT is that normally no one ever asserts that anything is red. This consequence is obtained with the help of another factual premiss: Normally when someone does what the naive person would call asserting that something is red, the "asserter" is not committing himself to some particular precise extension for "red." The remainder of the argument is, of course, parallel to that of the preceding paragraph.

I take it that there is an obvious reason why the MPT gets into these difficulties: Normally, if we affirm a sentence like S we do not thereby affirm some one of the propositions the MPT proposes. Instead, I should think we normally affirm *none* of those propositions, whatever they may be, but rather a different proposition, namely, the proposition that *the chip is red* — which is roughly the proposition that the chip possesses one or more color properties falling within a certain loosely delimited range of such properties — a range of properties we normally call by the adjective "red."

Further difficulties with the MPT, not dependent on the relation of propositions to assertion or belief, are not hard to generate. Suppose, for example, that we take our concern over color chips to be concern simply about the appearance of the chips to the unaided eye under normal observation conditions. (I.e., wavelengths or other "hidden" properties are set aside as irrelevant at the outset.) It is then possible to construct a series of reddish or red color chips, each indistinguishable in color appearance from its immediate neighbors in the series, but each distinguishably more or less orange than chips farther away in the series. The series begins with chips that are clearly red and ends with chips that are clearly orange, and not red. Let the following argument refer to the chips in such a series. (Premisses are labeled with 'P'; the conclusion with 'C'.)

$P1$: Chip #1 is red.
$P2$: Chip #2 is indistinguishable from Chip #1 in color appearance.
$P3$: If any chip c is indistinguishable from a red chip in color appearance, c is a red chip.
C: Chip #2 is red.

Clearly, by iteration on this pattern, we can arrive at the false conclusion that the orange chips are red. In the view of the MPT such an argument is clearly valid, so long as "red" is univocal; and $P1$ and $P2$ are true. So, to explain why the whole series of arguments is not sound, attention must focus on $P3$. The theory will have to claim that if "red" is univocal in $P3$, $P3$ is just simply false.

Here we have another untoward consequence of the MPT, for $P3$ is quite true, given the notion of red with which we began the argument, provided simply that the color appearances which determine whether or not a thing is red are the usual ones, and not some arbitrarily truncated set. This difficulty about the MPT approach to the series of arguments discussed just now is not restricted to arguments in which color appearance plays the crucial role. A similar argument can be constructed utilizing wavelength as the defining characteristic of redness. In each case, the same problems arise: The MPT is forced to draw arbitrary lines around various extensions of "red" and, as a result, it claims that the premiss which plays the role $P3$ plays above is just false. But these premisses are in fact all quite true, when normal notions of red are employed. One cannot legitimately escape the paradoxes generated by these arguments in the way the MPT escapes them. (The basis of the proper way to escape the paradoxes will be discussed briefly in Sect. II. It depends upon the failure of bivalence for $P3$ — $P3$ is quite true, but not completely true, so that each member of the series of arguments generating the paradox is not completely sound. Precise sense needs to be made of this, and Sect. II lays the groundwork for so doing. The final conclusion of a chain of not quite sound arguments can be false, in much the same way that the conjunction of a series of highly probable stochastically independent events can itself be highly improbable.)

The demise of the MPT means that all attempts to save precision and its accomplice, bivalence, have expired. All these theories stem from a certain false view about the relation between language and the rest of the world. If we can get clearer about the real nature of that relation, perhaps we can avoid the difficulties which have plagued the theories discarded above.

Let the predicate letter '*R*' abbreviate "is red"; let the individual constant '*a*' name the chip we struggle with. Is '*Ra*' true? To answer this question, standard model theory assigns either a property or a set to the predicate letter '*R*'. Assume, for simplicity in what follows, that the model theory we are using assigns properties, rather than sets, to predicate letters. Then, if the chip has the property assigned to '*R*', "*Ra*" is true, according to the received view. Note, however, that it is here assumed that the individual particulars have their properties (and stand in their relations) no matter which predicate letters, if any, are assigned to these properties or relations. That is, the properties of a thing are language-independent.[5]

It is also part of the received view to hold a sort of law of the excluded middle for property instantiation: Any given particular either has or does not have any given property (and it either stands or does not stand in any given relation to any given particulars). It seems to me that this law is true, and that models constructed according to it are acceptable.[6] However, if we accept this law in addition to the standard mechanisms outlined above, it turns out that every meaningful predication is either simply true or simply false. Precision is guaranteed by the straightforward way in which predicates are assigned to properties. Herein lies the problem with the received theory. It assigns properties and relations to predicates in an artificially precise way. It is by no means the only way we might sensibly go about giving truth conditions for atomic sentences in a first-order predicate calculus.

Note that the received view, as characterized above, does not include an explanation of *why* meaningful predicates must signify a definite property. Even if the predicate expressions of English were by some miracle all perfectly precise, we could still ask about the possibility of having predicate expressions in, say, an artificial language which were not precise. Bivalence as a principle of logic would fare poorly even then, unless we could show that for some reason each predicate expression *must* stand in relation to the world in the way the received doctrine describes. I can think of only two reasonable ways that one might try to show that: (1) One might claim that the language individuates the properties—i.e., that the properties possessed by a particular are not ever language-independent, so that there is an automatic correspondence between predicate expressions and properties. Properties, for example, may be thought of as the *meanings* of the predicate expressions. (2) Instead of making claim (1), or by way of expanding upon claim (1), one might claim that no explanation of truth conditions for predication, other than the traditional one, can be coherently formulated. Sect. II of this paper consists of a reply to claim (2). We may concentrate on claim (1) now.

The reasons one might adduce to support claim (1) go quite deep. There are notorious difficulties over the individuation of properties, and one way to begin to solve them is to adopt the stance that admissible properties are linguistically or conceptually determined, with the result that what counts as one property for members of one community need not count as one property for members of some other, possibly hypothetical, community. This way of looking at properties is at least verbally at odds with the "received" view characterized earlier; so that one who adopts this new way of talking about properties may think that I am right to attack the received view, but that I am wrong in locating the proper place to attack. According to him bivalence is saved by dropping part of the received view.

But no matter what one says about properties and their dependence upon language, or lack thereof, the point of contention between the received view and my view is unresolved. For, I am claiming that it is artificial and misleading to require the *extension* of each predicate to be precisely determined when doing formal semantics, because it is entirely possible to have meaningful predicate expressions whose extensions are only partially determined. Nothing has been said so far in the received view or in the new theory of property individuation to indicate why we must ignore this possibility.

---

[5] There are presumably some relational properties which are not thus language independent — e.g., the property of being called by the word "green" in English — but these are of no special consequence in the discussion.

[6] Even if properties admit of degrees of instantiation, this principle causes no trouble.

In particular, in the case of natural languages, even if we assume that properties are individuated by language somehow, surely there remains *some* language-independent (perhaps indescribable) factual basis for ascribing properties to objects. So long as properties are not the ultimate language-independent grounds for the truth or falsity of predication — whether or not properties are thought of as meanings of predicates — the question will arise: What reason is there for supposing that, given an object and a property, the raw non-linguistic facts about the object determine that the object either has or does not have the property?

Hence, my argument in favor of the contention that "red" in English is imprecise amounts to this: On the surface, "red" appears to be imprecise, on the basis of commonly experienced problems in deciding its application. Moreover, all reasonable attempts to explain away this appearance fail, for one reason or another, since there is no reason to suppose that meaningful predicates *must* have completely determinate extensions,[7] and there is every reason to suppose that in the development of natural languages no straightforward connection between predicate expressions and the ultimate factual grounds of truth and falsity is maintained. Instead of holding to the principle of bivalence, for the sake of expediency in logic, it would seem to me more reasonable to explore ways of handling propositions which are neither true nor false in a logical way.

II.

It is time now to fulfill promises made earlier. The goal now is to show that a coherent account of imprecise predication can be given. To do this, I shall describe a method, adapted from Zadeh and Goguen, for giving formal interpretations for predicates in the first-order predicate calculus;[8] this method can be used to make the predicates either precise or imprecise, as desired. Once the method for interpreting predicate letters is fixed, rules for deciding the truth values of the sentences of the calculus can be given, if we assume the interpretation of individual constants and operation letters is done in the usual way. However, rules of truth will not be discussed in detail here.

Let the language to be interpreted be the first-order predicate calculus, with a denumerable number of predicate letters for each finite number of places, and a denumerable number of individual constants. We restrict our attention to the interpretation of atomic sentences consisting of an n-place predicate letter followed by n individual constants, where not all the constants need be different, and $n \geq 1$. An interpretation M of these sentences consists of the following:

(1) A non-empty set D, called the *domain* of the interpretation.

(2) A set I, called the *index* of the interpretation, having at least three members.

(3) A set E, called the *set of possible extensions*, consisting of all the ordered pairs whose first members are n-place predicate letters ($n \geq 1$) and whose second members are ordered n-tuples of elements of D, where the number of places of the predicate letter must equal the number of places in the n-tuple in each case.

(4) A finite set F of predicate interpretation functions, each of which has some subset of E as domain and some subset (perhaps not a proper subset) of I as range. To insure that every predicate letter gets at least a minimal interpretation, we also require that for each predicate letter, there exist some member of F which has in its domain a member of E containing that predicate letter as first member.

(5) A function d, called the *denotation function*, which assigns to each individual constant an element of D. For simplicity of notation, we shall use 'a' in those places where 'd('a')' would be strictly more accurate. I.e., the assignment of denotations to individual constants will be tacitly assumed, without notational reminders.

Intuitively, any given interpretation function tells us for various ordered n-tuples of elements of D whether or not they fall within the extensions of various n-place predicates. An interpretation function does this by assigning to a paired n-place predicate and n-tuple some element of I. Suppose for convenience that I has exactly three elements, which we name "In," "Out," and "Borderline." (Discussion of what happens when I is allowed to expand will be held until later.) Then, if some interpretation function f has the value In for $\langle R, \langle a, b\rangle\rangle$ as argument, $\langle a, b\rangle$ is *in* the extension of 'R' as far as f is concerned; whereas if f assigns Borderline to $\langle R, \langle a, c\rangle\rangle$, f doesn't decide the question whether $\langle a, c\rangle$ is in the extension of 'R'.

---

[7] In view of the fact that I have not yet showed how a coherent account of truth conditions for imprecise predicates can be given, my summary argument here must be construed partly as a promise, to be satisfied at least to some extent, I hope, by Sect. II.

[8] Cf. L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8 (1965), pp. 338–353. For an especially valuable further development of these ideas in relation to a deductive calculus, see Joseph A. Goguen, Jr., "Categories of Fuzzy Sets" (doctoral dissertation), (Ann Arbor, 1969), and "Logic of Inexact Concepts," *Synthese*, vol. 19 (1968–69), pp. 325–373.

No requirement of consistency is placed on the set of interpretation functions — i.e., one member of F may assign In to $\langle R, \langle a, b\rangle\rangle$ while another assigns Out to $\langle R, \langle a, b\rangle\rangle$. This is a case of conflict, in which the semantical rules do not give a consistent answer to the question whether $\langle a, b\rangle$ is in or out of the extension of 'R'. It should not be thought, however, that we should then say that according to the interpretation under consideration, $\langle a, b\rangle$ is both in and out of the extension of 'R'. Rather, we should say that, in the given interpretation, which of course includes *all* the interpretation functions, the semantical rules conflict, with the result that $\langle a, b\rangle$ is neither in nor out of the extension of 'R': The truth of applying 'R' to $\langle a, b\rangle$ is in doubt. We shall want a rule of truth which tells us that in such circumstances "$R(a, b)$" is neither simply true nor simply false, presumably. We do not make a shambles out of the concept of truth by claiming that "$R(a, b)$" is *both* (simply) true and (simply) false. Our interpretations need not and will not sanction the derivation of such contradictions.

In general, the sentence formed by writing an n-place predicate letter 'G' followed by n individual constants "$a_1, a_2, \ldots, a_n$" will be *true* under a given interpretation M if and only if (1) some interpretation function in M has In as value for $\langle G, \langle a_1, a_2, \ldots, a_n\rangle\rangle$ as argument, *and* (2) no interpretation function in M has Out as value for this argument. (Thus, if one member of F assigns In to this argument, while another assigns Borderline to it, the truth value of "$G(a_1, a_2, \ldots, a_n)$" remains truth.) This same sentence will be *false* if and only if (1) some interpretation function in M assigns Out to $\langle G, \langle a_1, a_2, \ldots, a_n\rangle\rangle$ *and* (2) no interpretation function in M assigns In to this argument. If, and only if, "$G(a_1, a_2, \ldots, a_n)$" satisfies neither the conditions laid down here for being true nor the conditions for being false, it is neither true nor false. It will prove more convenient later if we think of the sentence as having a third truth value in this case; perhaps we can call it Neutral.

The possibility of having conflicting interpretation functions allows us to construct a formal structure which captures the features of a scientific language formerly sometimes described by means of reduction pairs. In our system we do not "introduce" predicates by means of "meaning postulates" some of which might be reduction sentences, as Carnap suggested.[9] For, if we did, it might be supposed that the truth or falsity of the postulates under a given interpretation should be determined in the traditional way. Rather, I suggest that the semantical rules be included in the structure of the interpretation functions themselves, where it is apparent that they are a part of the semantical mechanism for the interpretation of the language. In this way we avoid the problems associated with convergent pairs of reduction sentences which have consequences which are open to empirical falsification.

For example, if we try to adopt the reduction pair

$$(\forall x)\ (T_1x \supset (R_1x \supset Cx)\ ) \qquad (1)$$

and

$$(\forall x)\ (T_2x \supset (R_2x \supset \sim Cx)\ ) \qquad (2)$$

in the usual logic, as a set of meaning postulates on 'C', we can deduce

$$(\forall x) \sim (T_1x \& T_2x \& R_1x \& R_2x) \qquad (3)$$

which is presumably empirically testable if T_1 and T_2 are tests one might perform on some object and R_1 and R_2 are possible results of the tests. Various remedies for this malady have been proposed but my purpose here is not to explore the merits of these; rather I wish to point out how the features of scientific theory which the original reduction pair was designed to explicate can be formalized within the model theory proposed in this paper.

My suggestion is, roughly, that (1) and (2) be replaced by two interpretation functions for 'C' — one to do the job (1) was supposed to do, and one to do the job of (2). Presumably, these interpretation functions would be coordinated with interpretation functions for 'T_1', 'T_2', 'R_1' and 'R_2' in such a way as to make sentences *something* like (1) and (2) true under these interpretations. Since I have not given rules of truth for quantified sentences, nor for compound sentences, the discussion remains on the intuitive level here, but it is not too difficult to see how my account might be filled out in detail.[10] However, this much is clear: Whatever would take the place of (1) and (2) would not have (3) as a logical consequence, because the applicability of 'T_1' and 'R_1' to an object would guarantee the applicability of 'C' to that object only if no other rule for interpreting 'C' conflicted. Thus, suppose some object a is such that all the members of F assign In to $\langle T_1, a\rangle$, $\langle T_2, a\rangle$, $(R_1, a\rangle$, and $\langle R_2, a\rangle$. If the two interpretation functions for 'C' have been set up correctly, one of them assigns In to $\langle C, a\rangle$ and one assigns Out to $\langle C, a\rangle$. The result is that the truth value of "Ca" is Neutral, since the interpretation functions conflict. But the warrant for deriving (3) from (1) and (2) under the old system was that unless (3) were true, both "Ca" and "$\sim Ca$" would be true. That warrant is now removed.

---

[9] Rudolf Carnap, "Meaning Postulates," *Philosophical Studies*, vol. 3 (1952), pp. 65–73.

Thus, a scientific theory with such rules of interpretation for the term "C" would be unable to say either that the object named by 'a' is a C or that it isn't. This inability to decide the question might be grounds for revising the theory, but it is not grounds for saying the theory is self-contradictory, or that it contradicts the empirical evidence.

Aside from cases of conflict between the rules for applying predicates, we noted earlier that an interpretation function may have any subset of E as domain. This means that a given function f may not provide any interpretation at all for a given predicate letter, and that f, even if it does provide some interpretation for a given predicate letter, may still not say with respect to *every* object in the domain whether or not that object is in or out of the extension of that predicate letter. This allows another sort of imprecision to arise. In such cases, we may say that there are gaps in the information provided by the interpretation function f. And although the restrictions on M require that every predicate letter receive at least some interpretation by at least one member of F, it is possible that the gaps in information provided by all the members of F should overlap in such a way that for a given predicate letter the set F says nothing at all about its applicability to a given member of D. In such a case, the rules of truth given above would assign the truth value Neutral to the sentence.

An example of the concrete application of this aspect of our theory derives once again from semantical theories which employ meaning postulates. If the meaning postulates (1) and (2) given earlier are the only postulates for 'C', then presumably a consistent theory (interpreted traditionally) containing these postulates is silent with respect to the question whether some object a such that $\sim R_1a$ and $\sim R_2a$ has C. It is this sort of feature which our method of interpreting predicate letters can capture, without resorting to placing meaning postulates in the object language itself, by merely arranging the interpretation functions properly.

Finally, we turn our attention to the third way in which imprecision can arise: The relevant interpretation function fails to place an ordered n-tuple of elements of D either in or out of the extension of some n-place predicate letter; instead, the function places the n-tuple on the borderline of the extension by assigning Borderline to the appropriate ordered pair. Thus, assuming that the interpretation function g is the only predicate interpretation function in F which has $\langle P, \langle a, b\rangle\rangle$ in its domain, if $g(\langle P, \langle a, b\rangle\rangle) =$ Borderline, the rules of truth tell us that "Pab" has the truth value Neutral.

What sort of concrete case is this formal structure intended to picture? Cases like the "red" situation described in Sect. I. These cases are among the most interesting and the most difficult to handle of the cases in which imprecision of predicates may be manifested. I also suspect this sort of case is by far the most numerous in actual occurrence within natural language use.

It seems to me that several things may be going on in such cases — the details of which the formal theory of interpretations proposed here does not depict, but with which the formal theory nevertheless is compatible. For example, I think that the problem about "red" arose because there really is no one (determinate or determinable) property of being red, which could serve as the ultimate factual basis for grounding the truth or falsity of the sentence S. Rather, there are many closely related color properties which from our point of view arrange themselves into a kind of natural sequence along the various dimensions of hue, saturation, and brightness. These color properties are each completely determinate, and any given object either has one of them (at a particular time, on a particular part of its surface) or it doesn't. But the word "red" does not signify some precisely bounded set of these properties. If S is symbolized by "Ra," and the color chip is placed in D, we can obviously arrange our interpretation functions for 'R' to give this result.

---

[10] Cf., for example, John Rosser and Atwell Turquette, *Many-Valued Logics* (Amsterdam, 1952), where consideration is given to calculi with three and more values. I restrict my attention in this paper to the question whether predicates can have fuzzy extensions.

Not all the cases in which we should want an interpretation function to assign Borderline to an ordered pair in E are analogous to the case of "red," though. Consider, for example, the question whether a bicycle, or a tricycle, is a vehicle.[11] Here there is no obvious sequence of properties falling into some "natural" series which we can call to mind to explain why we may have trouble drawing a non-artificial line, as in the case of "red." Here, instead, it appears that there are various factors to be weighed and considered none of which are decisive in determining whether the tricycle is a vehicle. The result is that the sentence, "This is a vehicle," said appropriately of a tricycle, or a bicycle perhaps, is neither true nor false.

Of course the formal theory proposed in this paper does not provide any mechanism for mirroring the relevance of various weighted factors in determining the applicability of a given predicate letter to a given object in D. In the formal theory we must satisfy ourselves with assigning the appropriate element of I to elements of E, leaving out of account any explanation of the appropriateness. In so doing, we are, as it were, constructing only that function from E to I which represents the resultant of the counterbalancing forces acting on the relation between "vehicle" and the toy. I envision a fuller account of semantics which includes within the interpretation functions such matters as this weighting of factors.

In this regard, someone may quite reasonably object that if the description just given of the relation of "red" to actual properties in the world is correct, it is hardly appropriate to draw a sharp line between those color shades that are clearly red ones and those that are borderline cases of red ones. For this reason, it is no doubt better to have a more complicated structure in I. We could, for example, let I be the unit interval of real numbers 0 to 1. Then the value of a given interpretation function for a given element ϕ of E would represent the degree to which the second element of ϕ falls within the extension of the first element of ϕ as far as that particular interpretation function is concerned. The rules of truth and falsity would be changed accordingly, to give us a calculus somewhat parallel in structure to a probability calculus, with an infinite number of truth values. (Or, at least, they could be so changed.)

Thus, I leave open the question of how much structure I must have. What one chooses to do with I is partly a matter of how closely one wants to parallel the semantic structure of natural languages, and partly a matter of what will work to give us a manageable theory. Recent results indicate that one does not need so much structure in I as we get with the unit interval: Given various restrictions on I, one can define in a natural way various operations, such as union and intersection, on the fuzzy sets that represent the extensions of the imprecise predicates indexed in I. These operations provide a basis for giving rules of truth for compound sentences containing occurrences of imprecise predicates.

Of course, one can make some, or even all, of the predicates in the language being interpreted by M perfectly precise, by constructing the interpretation functions appropriately. I have not argued that we must interpret every predicate so as to make it imprecise; rather, I have argued that one can make the claim that predicates need not be precise into a coherent (and workable) one. I have argued for that because the construction of a precise account of imprecision is part of the argument that predicates can be (and are) imprecise, and that the principle of bivalence is not a logical truth.

Indiana University

Received July 16, 1971

---

[11] This problem was pointed out by H. L. A. Hart, *The Concept of Law* (London, 1961), p. 123.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 3, July 1972

III. ON AN ANALYTIC-SYNTHETIC DISTINCTION

BRUCE AUNE

SO much has been written on the analytic-synthetic distinction that it may seem pointless to produce still another essay on the subject. Yet not every approach to the distinction has been equally well explored. One neglected approach is suggested by some of C. S Peirce's remarks on meaning. I shall argue that an analytic-synthetic distinction drawn in the spirit of Peirce's pragmatism makes a significant contribution to critical philosophy.

As even Quine has admitted, a sharp distinction between the analytic and the synthetic is not difficult to draw if "analytic" is defined very narrowly.[1] The limits of what is purely logical are subject to dispute, but the basic subject of first-order quantification theory is well enough understood so that little objection is likely to arise if we specify a narrow sense of analyticity call it "L-truth," as encompassing all substitution instances (and translations thereof) of truths belonging to this basic subject. Trouble generally enters the picture when we go beyond this and attempt to specify a broader sense of analyticity that presumably includes "Princes are royal" and "Whatever is colored is spread out in space."

In attempting to improve upon Kant's vague and restrictive approach to this broader sense of the term philosophers have been inclined to follow Frege and argue that a statement is analytic in the broad sense if and only if it is either logically true or reducible to a logical truth by the help of definitions for the nonlogical expressions occurring in it.[2] Thus, if "A prince is a royal son" is a satisfactory definition, then the statement "Princes are royal" may be reduced to the logical truth "Royal sons are royal" and be declared, therefore, broadly analytic.[3]

This approach to analyticity has some very obvious merits. Most important, it shows just why a statement that is broadly analytic should be accepted as true. If a statement is a logical truth, it is clearly true (for whatever reason logical truths are true[4]); and if it is reducible to a logical truth—in the sense that, given certain definitions, we can demonstrate its equivalence to a logical truth by a series of logically valid inferences—then it is clearly true as well. This advantage is not negligible, since it takes us well beyond the stage of speaking vaguely of whether a predicate is contained in the concept of a subject or whether we can *see*, in some mysterious fashion, that a statement not L-true is yet true *a priori*.

In spite of these patent advantages Frege's approach unfortunately possesses some well known drawbacks. Perhaps the most familiar difficulty concerns the proper definition of a nonlogical expression. If such a definition is what Quine calls "discursive," its acceptability is based on a presumed synonymy (in some sense) of definiens and definiendum.[5] But this relevant sense of synonymy is notoriously obscure. Few, if any, words have exactly the same meaning — in a generic sense of "meaning." The locution "May I present my sister" plainly differs in import from "May I present my female sibling," even though philosophers commonly regard "sister" and "female sibling" as prime examples of synonyms. Obviously, there are many modes of meaning, and if all modes were to be included under "synonymy," probably no expressions could be allowed as proper synonyms. If we are to maintain that a discursive definition requires that definiens and definiendum be synonymous, we must therefore restrict the relevant kind of synonymy so that it covers only special modes of meaning; we might say that we are concerned only with *conceptual* synonymy. Unfortunately, little success seems to have been achieved in explicating this restricted kind of synonymy.

---

[1] See W. V. O. Quine, "Carnap and Logical Truth," in *The Ways of Paradox* (New York, 1966), pp. 100–123.

[2] See Gottlob Frege, *The Foundations of Arithmetic*, tr. by J. L. Austin (New York, 1966), pp. 3f.

[3] I am here using the word "statement" to mean "sentence understood according to a fixed interpretation." A more exotic usage is defended in my paper "Statements and Propositions," *Nous*, vol. 1 (1967), pp. 215–230.

[4] For purposes of this paper, I shall simply assume that a rationalist interpretation of logical truth is untenable. I have criticized this approach in my book, *Rationalism, Empiricism, and Pragmatism* (New York, 1970), ch. IV.

[5] See Quine, *op. cit.*, p. 112.

But even if this difficulty concerning synonymy can be resolved, Frege's approach has the additional defect of being excessively restrictive, since many truths that seem obvious candidates for the title "broadly analytic" involve descriptive expressions that are not explicitly definable. A familiar example is "Anything colored red is spread out in space." For any foe of rationalism, this statement seems analytic if anything does; yet its truth hinges on the meaning of "red," an expression apparently not capable of explicit definition. If truths of this kind are indeed analytic, then Frege's approach is insufficient on the ground that it cannot handle the entire field of analytic truths.

If a sufficiently broad distinction between the analytic and the synthetic is to be tenable, the difficulties with Frege's approach must in some way be surmounted. Specifically, the mode of meaning we might call "conceptual" must be explicated and a case must somehow be made for the analyticity of certain statements containing expressions not considered explicitly definable.

A promising strategy for accomplishing these tasks is suggested, I believe, by Peirce's discussion of the meaning of what he calls an "intellectual conception." His general view, roughly, was that the relevant meaning of such a conception is determined by special inferences we may draw from statements involving the conception. Since so-called ostensively defined predicates warrant no special inferences, it is clear that his view (vaguely as I have described it) is incompatible with a key doctrine of traditional empiricism. As a first step toward working out the analytic-synthetic distinction that is suggested by Peirce's views on meaning, I shall offer some brief critical remarks on the notion of an ostensive definition.

To attempt an ostensive definition is to try to specify an expression's intension solely by reference to a restricted part of its extension. But such an attempt cannot be expected to succeed. Observe, to begin with, that any group of entities one examines, however large it may be, will be characterized by coextensive predicates. If *A* is a class of such entities, then, no matter how large it may be, there are nonsynonymous predicates *P* and *Q* such that for any *x* in *A*, *P* is true of *x* just in case *Q* is true of it. This fact may be proved by reference to Goodman-type predicates such as "is *F* while *S*" (where "*S*" is any true statement) but it should be obvious to anyone familiar with the large literature on linguistic relativity.[6]

The significance of this fact for the subject of ostension is easy to see. If the objects of any class of entities one examines are characterized by coextensive predicates, then there can be nothing in the narrow confines of the ostensive situation that will fix the meanings of these different predicates. To put the point in an old-fashioned way, the word one is interested in may designate any one of the potentially infinite variety of properties possessed by all of the objects one has examined, and there is nothing in the ostensive situation that can single out the particular property one has in mind. This is not merely to say that a so-called ostensive definition can always be misunderstood; it is to say that an ostensive gesture (or a group of such) cannot possibly pin down the distinctive meaning of a predicate and cannot, therefore, provide a definition in any reasonable sense of the word. A man may, it is true, form the right idea of a predicate's meaning by observing such a gesture, but the idea he forms is not thereby shown to be correct: it has the status of a mere hypothesis. Countless other interpretations will accord equally well with the gesture (or gestures) observed, and to determine that the idea formed is actually correct, an additional procedure is required.

If we want to shed light on a predicate's distinctive meaning by reference to a class of entities to which it truly applies, the class will have to include possibilities as well as actualities. An explicit reference to possibilities will no doubt be unacceptable to many philosophers, but essentially the same result can be attained if we attend to a class of descriptions we may easily construct. Supposing that all actual *F*'s are also *G*'s and vice-versa, we might proceed to construct unusual descriptions of *F*'s and *G*'s and then consider whether the *F*'s as so described would count as *G*'s and vice-versa.

As a concrete example of this procedure, consider the predicates "rational animal" and "biped capable of laughing." Since these predicates are (as we may assume) coextensive, we cannot point to something that is correctly described by one but not by the other. We can, however, conceive of such a thing; we can conceive, for example, of a very stupid, patently nonrational laughing gorilla. The statement "If x is a stupid (nonrational) laughing gorilla, then x is a laughing biped but not a rational animal" would seem to be true by virtue of the predicates involved. Assuming that the antecedent of this statement is consistent, it would then follow that the predicates "rational animal" and "biped capable of laughing" are not necessarily coextensive and not, therefore, synonymous in a narrow conceptual sense.

---

[6] If '*S*' is a true statement, then the predicate "is *F* while *S*" is coextensive with "is *F*." For further discussion and less artificial examples, see my *Rationalism, Empiricism, and Pragmatism*, *op. cit.*, pp. 162–166.

This general strategy for casting light on the distinctive meanings of coextensive predicates is fully in line with a standard practice of contemporary analytic philosophers—that of clarifying concepts by devising thought-experiments. The strategy also recalls Peirce's famous claim that the meaning of an intellectual conception relates to "would-be's" rather than to what merely is.[7] Put in today's language, his idea is that to know what a predicate means is not to know what it actually applies to; it is to know what it *would* apply to *if* that or that were the case. As far as learning is concerned, the idea would be that to learn a predicate's meaning is to learn to draw appropriate inferences from statements in which that predicate occurs.[8]

Some readers will no doubt feel acutely nervous in seeing the subjunctive mood used so freely, but I hope this nervousness can be suppressed for the time being. My aim at present is simply to introduce the basic Peircian theme that a predicate's conceptual meaning has to do with the conclusions one is entitled to draw from premisses involving that predicate. The required variations on this theme can be left until later.

A point deserving of immediate attention is that Peirce's claim leads directly to an explication of conceptual synonymy by a *via negativa*. For anyone at all tempted by Peirce's approach to meaning, the following claim should have a strong intuitive plausibility: If nothing follows from a statement S containing a predicate P that does not also follow from a statement that differs from S only in containing a predicate Q in place of P (and vice-versa), then the two statements and *a fortiori* the predicates P and Q do not differ in conceptual meaning. If it makes no difference in what we can infer to interchange the predicates P and Q in a statement, then the two predicates can hardly differ in conceptual content. As Peirce would say, a difference that makes no difference is really no difference at all.

Though the claim just made is, at least in my view, highly plausible, it is subject to numerous counter-instances. These counter-instances are of two fundamental kinds. First, nonsynonymous statements may yet be self-contradictory so that, given standard logics, whatever follows from the one will follow from the other: everything follows from a contradiction. Secondly, some nonsynonymous predicates are necessarily (or demonstrably) coextensive so that whatever follows from an open statement containing just the one predicate will also follow from a corresponding statement containing the other. This last example is related to the principle that if P entails Q when P and R are demonstrably equivalent, then R entails Q.

Can these counter-examples be avoided so that the claim made earlier can be saved? I think "Yes." To avoid the first group of counter-examples it is sufficient to generalize the claim along the following lines: If nothing follows from *any* extensional statement S containing a predicate P that does not also follow from a statement S^* that differs from S only containing Q in place of P (and vice-versa), then P and Q do not differ in conceptual meaning. This amounts to saying that F and G are cognitively synonymous when all pairs of extensional statements differing only with respect to these predicates have the same implications. Since not all pairs of statements differing in this way are self-contradictory, this generalized claim avoids counter-instances of the first kind.

The second kind of counter-example is more difficult to avoid. One way of approaching it is to employ a distinction between the mediate and the immediate consequences of a statement. This distinction may be illustrated by the predicates "triangular" and "trilateral." In a pure geometry these predicates are demonstrably coextensive so that, for any statement P, "a is triangular" implies P if and only if "a is trilateral" implies it. It seems obvious, nevertheless, that these predicates differ importantly in conceptual meaning. Is there any real basis for this idea? I think there is. The necessary coextensiveness of "triangular" and "trilateral" is not a matter of basic principle; it is a derivative truth that requires a demonstration.

---

[7] See Charles Hartshorne and Paul Weiss (eds.), *The Collected Papers of Charles Sanders Peirce*, Vol. V (Cambridge, Mass., 1934), sect. 5.464–58.

[8] This theme has been stressed by Wilfrid Sellars; see his "Is There a Synthetic A Priori?" in Sellars, *Science, Perception, and Reality* (New York, 1963), pp. 298–320

From "*a* is triangular" we can immediately infer "*a* has three angles," but to infer this from "*a* is trilateral" we need some extra premisses (or we must employ a derivative rule). These extra premisses may be axioms, but they are extra premisses for all that. An axiom is not a redundancy.

If a predicate's conceptual meaning is to be explicated by the *immediate* consequences of extensional statements in which it occurs, the rules that warrant these consequences must possess two distinct features. First, they must be *primitive* rules; they must not, that is, be mere consequences of other rules and axioms. This is needed if we are to distinguish the meaning of predicates like "triangular" and "trilateral." If a consequence *C* is derivable from the conjunction of *P* and certain axioms, then an immediate inference from *P* to *C* is derivatively valid; we may employ the axioms to validate the inference from *P* to *C*. Secondly, the rules in question must be "material" as opposed to formal rules of inference.[9] The latter tell you what follows from statements in virtue of their logical form; they do not tell you what follows from statements in virtue of their nonformal content or meaning.

If it is objected that the notion of a material rule of inference is excessively obscure, an appropriate reply is that formal and material rules of inference have approximately the same status. In fact, the distinction between them (and thus between a statement's form and content) is no sharper than the distinction between a logical and a nonlogical expression, which different philosophers draw in different ways. Carnap, for example, once classified[10] "is a member of" as a logical expression, while Kneale[11] regards it as distinctively mathematical. In my view neither writer can be said to be *correct* in this matter. A sharp distinction between the logical and the nonlogical can no doubt be drawn, but it is bound to be arbitrary in a significant degree. The same is true of a distinction between formal and material rules of inference.

Thus far I have been speaking of simple predicates like "triangular" rather than compound ones like "red and rough." To accommodate both kinds of predicates in a definition of synonymy, a recursive procedure is appropriate. The following seems reasonable in view of my preceding remarks:

> Two predicates *F* and *G* are conceptually synonymous (do not differ in conceptual meaning) if and only if either (a) *F* or *G* is a simple predicate and for any extensional statements *S*, *S**, and *C*, if *S* differs from *S** only in containing *F* at one or more places where *S** contains *G*, then *C* follows immediately from *S* by a primitive rule of material inference if and only if *C* follows from *S** by such a rule, or (b) *F* and *G* are compound predicates and if *F* differs from *G* in containing a predicate *Q* at one or more places where *G* contains *R*, then *R* is conceptually synonymous with *Q*.[12]

This definition sounds complicated, but it is merely a development of the idea introduced in connection with Peirce, namely, that two predicates do not differ in conceptual meaning when the replacement of one for the other in a statement makes no difference for the inferences we may draw. I have obviously not explained all of the qualifications imported into my definition, but they raise no issues of philosophical interest; they are needed merely to avoid difficulties of formal detail.

It will be observed that, according to my definition, few predicates are likely to be exactly synonymous in a cognitive sense. This is not an entirely unwelcome consequence, however. Even in an intuitive sense few familiar predicates seem exactly synonymous; those identified as synonymous in dictionaries or in everyday speech generally possess subtle or even obvious differences of meaning. My explication deals with a very strict sense of conceptual synonymy and it is wholly compatible with the idea that predicates may be regarded as synonymous in looser (or different) senses of the word.

Although my explication of conceptual synonymy is intended to be relevant to the notion of analyticity, my immediate aim in formulating it was not to support Frege's approach but to throw needed light on the conceptual meaning of a predicate. When I criticized ostension theories I argued that a predicate's distinctive meaning can be at least partly explicated by appropriate rules of inference. My definition of synonymy shows that the full conceptual meaning of a predicate may be given by such rules. To see this suppose that a predicate *P* has an element of conceptual meaning that is not exhibited by a rule of inference. It is then possible that a predicate *Q* might lack this special element of meaning, be therefore not synonymous with *P*, and yet be such that every extensional statement *S* containing *P* has the same immediate consequences as a corresponding statement S^*, where S^* differs from *S* only in containing *Q* where *S* contains *P*. But this possibility is ruled out by my definition of "synonymy."

---

[9] Sellars also emphasizes the importance of material rules of inference; see *ibid.*

[10] In a seminar given at U.C.L.A. in 1957 Carnap remarked that the so-called *Aussonderung* Axiom of set theory looked like a logical axiom to him.

[11] See William and Martha Kneale, *The Development of Logic* (Oxford, 1961), pp. 741f.

[12] For purposes of this definition, the rule of repetition ("From *S* to infer *S*") must be counted a limiting case of a rule of material inference.

Observe that this definition applies only to predicates. Although singular terms may also be conceptually synonymous, a definition applicable to them is beyond the scope of this paper.

It should be understood that, in my view, logically equivalent statements such as "*P* & *Q*" and "*Q* & *P*" are *not* conceptually synonymous; their lack of synonymy is what makes logical rules such as commutation nontrivial.

My view of conceptual meaning may seem perverse. If, as it implies, knowing the meaning of one predicate requires knowing the meaning of other predicates, how could one come to know the meaning of any predicate at all? The solution to this problem does not require us to postulate basic, unanalyzable concepts; it merely requires us to acknowledge that coming to have a concept is a gradual process of learning the relation of verbal elements to one another and to reality. It is only when a *pattern* of verbal relationships is finally mastered that one has any concept at all.

As far as the application of concepts to reality is concerned, it is not necessary to accept the empiricist view that some predicates must be associated with nonverbally distinguished universals or kinds. We can say, instead, that a language is applied by virtue of the tendency of its speakers to respond with particular verbal expressions when their sense organs are stimulated in certain ways.

The stimulus-response character of this kind of application will permit us to say that the conceptual content of a predicate is wholly constituted by its relations to other predicates. Statements of the kind "The word 'pig' applies to pigs" are, of course, true, but they are empty.

How does this view of meaning relate to Peirce's claim that the meaning of an intellectual conception concerns would-be's? How does it relate to the claim that to know the meaning of a predicate is to know what *would* be the case *if* such and such? The answer to this seems easy. A rule of the form "From '*P*' one may infer '*Q*'" may be expressed in the material mode by saying "If it were the case that *P*, it would be the case that *Q*."[13] Peirce, at least, seemed to accept this way of putting things; he called such material mode conditionals "guiding principles" of acceptable forms of inference.

There is, of course, more than one way to skin a cat and also more than one way to specify the relation between rules of inference and Peircian conditionals. If for some reason one does not wish to speak of "expressing a rule in the material mode," one can say that Peircian conditionals are simply true statements that are immediately derivable from a corresponding rule by conditional proof.

I myself find it convenient to adopt the first approach, but I have no objection to some other strategy. To indicate my tolerance in this matter I shall sometimes speak of the Peircian conditional "corresponding" to a primitive rule of immediate inference.

The nervousness I ignored some time ago has no doubt sprung up again at my casual use, a paragraph back, of a subjunctive conditional. I must therefore try to show that this nervousness is unnecessary in the present context.

Although a great deal has been written about subjunctive conditionals, it is not generally appreciated that a wide variety of logical relations would naturally be expressed in the vernacular by subjunctive statements. The point, however, is obvious once one considers it. Even the problematic cases of subjunctive conditionals fall into importantly different types. Some such statements are supposed to formulate natural laws while others formulate conditional truths about individual things or persons. In general, about the only thing you can infer from the fact that a statement is formulated as a subjunctive conditional is that its truth-value is not a function of the falsity or truth of its antecedent or consequent. And in an age when intensional logics are flourishing it should be common knowledge that the variety of logically distinguishable conditionals is potentially endless.

As I construe it, a Peircian would-be is fundamentally different from the problematic kinds of subjunctive conditional discussed by Goodman and others.[14] The familiar problems with those conditionals are due to the bearing of collateral information on their truth-value. But this is not true of a Peircian would-be.

---

[13] A rule of inference may also be expressed at the same linguistic level as the corresponding Peircian conditional by a locution such as "From the premiss that *P* one may conclude that *Q*." This means of formulating rules of inference is helpful in meeting rationalist objections based on the idea that metalinguistic rules of inference mention or concern the expressions of a particular language whereas the corresponding Peircian conditionals do not.

Consider this fragment of a conditional: If Verdi were a Frenchman then — what? Would he not have written *La Traviata*, or would this opera have been written by a Frenchman? Our indecision hinges on the different items of background information at our disposal. If we consider that Verdi did write *La Traviata*, we are inclined to conclude that *La Traviata* would then have been written by a Frenchman. But since we know that the opera was written by an Italian, we are tempted to conclude that if Verdi were a Frenchman, he would not have written the opera. Our problem is thus due to the bearing of collateral information on the hypothesis we are considering. We have a problem because our background information prompts us to draw conflicting conclusions from the hypothesis.[15]

Similar problems do not arise for the subjunctive conditionals expressing Peircian would-be's. The reason for this is that the rules of inference explicating the meaning of a predicate are primitive rules of *immediate* inference: they tell us that a conclusion may be inferred from a premiss *without the aid of collateral information*. Corresponding to such rules we therefore have object-language conditionals formulating immediate consequences. Since the truth of such a conditional will not depend on the truth-value of its antecedent or consequent, we naturally formulate the conditional in English in the subjunctive mood. But this formulation does not imply that the statement is of the same problematic kind as the one concerning Verdi or as those concerning causal connections that have been discussed so often in the literature.

How might we formulate our Peircian conditionals if we wish to employ a more technical idiom? I think the familiar horseshoe will do, provided we remember that the relevant conditionals are to be understood as unconditionally assertable or "conceptually necessary." Such conditionals — for example, "$(\forall x)$ (x is trilateral $\supset$ x has three sides)" — will have the status of axioms or (to use Carnap's terminology) meaning postulates; and if this is kept firmly in mind, I can see no objection to formulating them as Carnap did.[16] If we use the subjunctive mood when expressing them in English, this is only to indicate that their acceptability does not depend on the truth-value of their antecedent or consequent.

I have now said enough to tackle the main task of my paper: What can we mean by an analytic statement? How might we explicate the concept of analyticity? Here I follow Carnap.[17] An analytically true statement is simply one whose truth is a consequence of the meaning rules of language in which it is expressed. This claim is not just an empty form of words given the preceding discussion, for I have argued that every significant predicate owes its conceptual meaning to material rules of inference, which may be expressed either as explicit rules telling us what we may infer or as conditional statements of the kind Carnap called "meaning postulates." Carnap's terminology for analytic truth is "*A*-truth" and this notion can be explicitly defined for our purposes as follows:

> A statement S is *A*-true in L if and only if S is logically implied by the meaning postulates of L or follows from the primitive rules of immediate inference associated with the expressions of L.

Since logical truths follow vacuously from anything whatever, we can allow that every logical truth is an analytic truth but not vice-versa.

Early in this paper I discussed Frege's approach to analyticity. The explication just offered differs from his, but it does so chiefly in making room for analytic statements containing essentially expressions that are not explicitly definable. Thus if predicates F and G are conceptually synonymous by my definition, this fact can be used to establish the analyticity in his sense of appropriate statements containing F or G. Of course, my definition of conceptual synonymy is very strict, and it might be, for example, that "prince" and "royal son" will fail my test even though "x is a prince if and only if x is a royal son" seems an obvious verbal truth. But this would not rule out a Frege-like reduction of "A prince is royal" to a logical truth. Instead of using the biconditional just mentioned as a definition, we could use it as a derivative analytical truth — one that, like "x is trilateral if and only if x is triangular," can be inferred from our meaning postulates.

---

[14] See Nelson Goodman, *Fact, Fiction, and Forecast*, 2nd ed. (New York, 1965), pp. 3–8, and Roderick Chisholm, "The Contrary-to-Fact Conditional," in Herbert Feigl and Wilfrid Sellars (eds.), *Readings in Philosophical Analysis* (New York, 1949), pp. 482–497.

[15] See Nicholas Rescher, "Belief-Contravening Suppositions," *Philosophical Review*, vol. 70 (1961), pp. 176–196.

[16] See Rudolf Carnap, *Meaning and Necessity*, 2nd ed. (Chicago, 1956), pp. 222–229.

[17] *Ibid.*

Assuming that the general strategy of my approach to analyticity is reasonably clear, I now want to add some remarks and qualifications that are prompted by the critical literature on the subject. Since my approach is on the surface similar to Carnap's, I shall begin by commenting on a standard criticism of his position.

It is often objected that Carnap's epithet "meaning postulate" does not really help to clarify the special status awarded to analytic truths.[18] Merely to label certain statements as meaning postulates is not to show us why those statements are true, why they deserve to be accepted, or how their presumed acceptability differs from that of truths of other kinds. All this must obviously be shown if the use of Carnap's label is to be worth the trouble of writing it down.

Although this objection is sound as far as it goes, I have implicitly indicated the kind of response I would make to it. Far from merely labeling certain conditionals as meaning postulates, I have laid down the essentials of a pragmatic theory of meaning and have introduced my postulates in terms of that theory. To relate my treatment to Peirce's thinking I need only cite his famous remark:

> In order to ascertain the meaning of an intellectual conception, one should consider what practical consequences might conceivably result by necessity from the truth of that conception; and the sum of these consequences will constitute the entire meaning of that conception.[19]

For Peirce, a "consequence" is not an observable effect, as is commonly supposed it is a conclusion properly drawn from a premiss.[20] I use Carnap's word "meaning postulate" to mark those conditionals whose antecedent contains the predicate (or "intellectual conception") that is partly explicated by the "consequence" appearing in the consequent. For me as for Peirce, the sum of these consequents gives the entire conceptual meaning of the predicate in question. Thus, although I use Carnap's terminology, I do not use it without explanation or justification. I doubt that Carnap would have objected to my use of his words.

A qualification I might make at this stage is that even though a certain meaning postulate is unconditionally assertable for a person using a predicate with a particular sense, the postulate is not for that reason immune to possible revision or outright rejection. Peirce held (rightly, I think) that it is part of the task of science to change old meanings and to develop new ones, sometimes on the basis of purely factual considerations. Since I agree with Peirce on this, I cannot deny that a meaning postulate acceptable at one time may require correction at another time. But this merely implies that if we wish to call our meaning postulates "necessary truths," we cannot conceive of such truths as something it would be irrational ever to abandon.

I might add in this connection that my approach to analyticity can accommodate some of the looser conceptual necessities discussed in connection with the notion of a criterion. A normic criterion in Scriven's sense is necessarily evidence for the presence of an appropriate characteristic, even though it does not insure the presence of that characteristic.[21] I can accommodate such criteria by allowing that Peircian conditionals may sometimes have a probabilistic force. Thus, "For any x, if Fx, then probably Gx" might be the general form of some Peircian conditionals. Such conditionals are obviously relevant to the explication of so-called open concepts.

A very serious objection to my approach should not be neglected, however; it is that it is highly questionable whether anything like meaning postulates is valid for an entire language. Is it really plausible to say that we could in principle draw up a list of meaning postulates valid for the whole of English? If Chomsky and others are right, something like this might possibly be done. They argue that the ability of a native speaker to understand countless English sentences not hitherto encountered shows that they must be operating with definite rules.[22] If such rules exist, at least some meaning postulates of the kind I have described may then be valid for native speakers. On the other hand, as Quine has observed, the existing experimental evidence does not indicate an "impressive uniformity" among native speakers in putting favored candidates of analytic truths into a special category.[23]

---

[18] See Quine, *loc. cit.*

[19] This remark is quoted by Amelie Rorty in *Pragmatic Philosophy* (Garden City, New York, 1966), p. 5. I cannot locate the original source in Peirce.

[20] See John Boler, *Charles Peirce and Scholastic Realism* (Seattle, 1963), pp. 98f. My interpretation of Peirce is unusual but, I believe, true to his mature views. When Peirce speaks of consequences as "practical," he does so to relate them to rational *conduct*. For him, a belief is a rule of action, but the relevant action is conceptual; it is the action of rational inference. The practical consequences that, he says, result by necessity from the truth of a conception are therefore the conclusions that would result in a rational mind from the supposition that the given conception is true. These conclusions would result by necessity in a rational mind in the sense that (or because) they are implied by the truth of that supposition.

[21] See Michael Scriven, "The Logic of Criteria," *Journal of Philosophy*, vol. 56 (1959), pp. 857–868. My approach to criteria is expounded in my book *Knowledge, Mind, and Nature* (New York, 1967), pp. 114f.

Fortunately, it is not necessary for me to take a stand on this issue. As far as I am concerned here, it may well be true that there are no meaning postulates valid for every speaker of English. When we speak of *the* meaning, or meanings, of an English word, we are operating at a very high level of abstraction; we are ignoring differences in dialects or even idiolects and concentrating on common ground. If this common ground were not substantial, individual speakers probably could not communicate, at least effectively. Yet this does not imply (even though it might be true) that there must be *full and uniform* agreement on significant aspects of meaning, on aspects that could be expressed by meaning postulates. It is entirely possible that the relevant common ground is due merely to a strong family resemblance (in Wittgenstein's sense) between the meanings different speakers attach to a given word. Such family resemblances could account for the social aspects of language learning, but they might be inadequate to justify the validity of certain meaning postulates for every speaker of a given language.

If, however, I am willing to express this kind of uncertainty about the validity of certain meaning postulates for a language like English, why do I insist on the importance of an analytic-synthetic distinction that requires such postulates? My answer may be expressed as a slogan: Determinateness of sense requires an analytic-synthetic distinction for a large class of statements. If a man uses a predicate with a clear, determinate sense, then he is using it according to clear, determinate rules, which may be formulated as meaning postulates. Roughly speaking, to know exactly what you mean by an atomic predicate P is to know exactly what Peircian consequences you are prepared to allow from the open sentence Px.[24] You may indeed use a predicate with several clear alternative meanings, but then you must allow several different consequences to be drawn from different uses of the function Px. The predicate's definite ambiguity is in fact determined by the different implications relevant to different uses of it.

In everyday life, of course, we rarely use language in this determinate way, and we are therefore rarely prepared to specify a list of rules that fully elucidate the meaning of a given predicate. Normally, our purposes are only vaguely conceived — a few highlights stand out in our minds — and the precision of our language corresponds to the precision of our purposes. We are not merely careless in using language that has, for us, a clear, determinate sense; our meanings are as ill-defined as our purposes are vague. In practice we are only as clear and precise as we have to be. Grunts and gestures are often as satisfactory as words.

Since our everyday speech is at least to some extent a Heraclitean flux, a sharp analytic-synthetic distinction is perhaps impossible to draw for an entire living language or even for an entire idiolect. But this does not prove that the idea of a sharp analytic-synthetic distinction is unimportant for philosophy. On the contrary, the idea is of immense philosophical importance; it is as important as the idea of determinate sense or meaning. The distinction may be only an ideal, but then reflection on ideals seems to be a distinctive task of philosophy. Moral excellence, beauty, and, if Peirce was right, even truth itself are ideals. If an analytic-synthetic distinction must be accorded a similar status, its "overwhelming unimportance" is *not* thereby assured.[25]

University of Massachusetts at Amherst

Received May 5, 1971

---

[22] See Noam Chomsky, "Current Issues in Linguistic Theory," in Jerry A. Fodor and Jerrold J. Katz (eds.), *The Structure of Language*(Englewood Cliffs, N.J., 1964), pp. 50–118. On page 17 of this essay Chomsky remarks: "It is not entirely obvious whether or where one can draw a natural bound between grammar and 'logical grammar' in the sense of Wittgenstein and the Oxford philosophers."

[23] See W. V. O. Quine, "On a Suggestion of Katz," *Journal of Philosophy*, vol. 64 (1967), pp. 52–54.

[24] This claim requires qualification since some atomic predicates may have immediate consequences only when conjoined with other predicates. Thus, the sole meaning postulate associated with the predicate P could conceivably be of the form $(Px \mathbin{\&} Qx). \supset Rx$. The definition of conceptual synonymy given in this paper is designed to accommodate such predicates.

[25] This paper was read at a number of universities, and I have profited greatly from the critical comments I have received. My colleagues at the University of Massachusetts (particularly Gary Matthews) have also given me very helpful criticism.

American Philosophical Quarterly
Volume 9, Number 3, July 1972

IV. DEPICTION, VISION, AND CONVENTION

PATRICK MAYNARD

I. Pictures as “Graphic Symbols”

SOME recent uses of symbol and language metaphors for the pictorial arts (witness titles like “The Visual Dialogue,”[1] “The Grammar of Drawing,”[2] and remarks such as “The relationship between the painting and what it represents is symbolic,” under other titles[3]) have had the good intention of restoring appreciation for representation itself — largely through efforts to rescue the concepts of representation from the competing model of “the imitation of appearances.” This paper attempts to identify, consolidate, and assess some main lines of argument for this approach, but only from the restricted standpoint of their bearing upon the old and continuing business of capturing the appearances of the natural world.

Language metaphors are of course not new to language about the arts. But where in the heyday of artist-expression theories those metaphors had a flavor of aesthetic “pragmatics,” their appeal today suits a taste for art’s syntax and semantics. For a parallel in theory to the recent post-painterly cooling on the topics of art as the artist’s action and self-realization we may compare the relative places accorded representation and expression in the earlier “linguistic” doctrines of Croce and Collingwood with those in the theories of Gombrich and Goodman. “Expression,” after all, was only one of several terms figuring in disputes about the “modern art” movement at the beginning of the century, but it became the focus for much of the subsequent theorizing: a peculiar situation, since a crisis of representation lay near the heart of the revolution. Possibly the assumption was that we already understood depiction (which, as everyone knew, artists had been doing all along), and that our efforts should be spent explaining expression (which, supposedly, was what artists had really been doing all along), and nonfigurative design (which some were beginning to do for the first time).

Many of course knew differently, and sensed misconceptions about the nature of representation on all sides of the controversy. Thus Meyer Schapiro had observed in 1937 how “in regarding representation as a facsimile of nature, the abstract artist has taken over the error of vulgar nineteenth century criticism. . . . If the older taste said, how exactly like the object, how beautiful!—the modern abstractionist says, how exactly like the object, how ugly!”[4] Latterly, the correction has been made by use of a symbol or language theory of representation. Writing about the modernist controversy some years after Schapiro, Leo Steinberg makes this case against the imitation view:

> The encroaching archaism of old photographs is only the latest instance of an endless succession in which every new mode of nature-representation eventually resigns its claim to co-identity with natural appearance. And if appearances are thus unstable in the human eye, their representation in art is not a matter of mechanical reproduction but of progressive revelation. . . . “Technical capacity in the imitation of nature” simply does not exist. What does exist is the skill of reproducing handy graphic symbols for natural appearances, of rendering familiar facts by set professional conventions.[5]

Where imitation becomes a boring technical feat it is not as imitation of nature but rather as imitation of other pictures.

I count three points in this on the topics of imitation, vision, and convention which have since received extensive elaboration. First, there is the principle that a picture could not give an imitation of *the* appearance of something since there is no such favored appearance. And in the “symbol” terminology there is a second, implicit, point that a picture does not yield even an imitation of *an* appearance of its subject,[6] since it is rather what various writers have termed "a symbolic cipher" for it, its "translation," or "structural equivalent in a given medium." (Some have urged further that a picture stands as such a symbol only for those who are acquainted with its style and medium "language.") Finally, in the remark that new kinds of "graphic symbols" reveal new appearances to us, there is the view that a cause of the instability of natural appearances is the instability of the representational conventions themselves.

---

[1] Nathan Knobler, *The Visual Dialogue* (New York, 1967).

[2] Colin Hayes, *The Grammar of Drawing for Artists and Designers* (New York, 1969).

[3] Frederick Gore, *Painting: Some Basic Principles* (London and New York, 1965), p. 36.

[4] “Nature of Abstract Art,” *Marxist Quarterly*, vol. 1 (1937), p. 86.

[5] “The Eye is a Part of the Mind,” *Partisan Review*, vol. 20 (1953); reprinted in S. Langer (ed.), *Reflections on Art* (Galaxy Book edition, New York, 1961): see p. 247.

Let us begin with the first of these observations. Everyone already knew that a given thing will have different appearances at different times, from different points of view, to different observers. What the principle asserts beyond this is that ideas about what we might call a representative or typical appearance of a thing may vary widely. Given a group of photographs of a house or a friend we are usually able to sort among them for good likeness. Yet if any of those images presents what can be called an imitation of an appearance, they all do, and since it is therefore the appearances themselves which are being graded for correctness, something besides close imitation of an appearance must be involved in the criteria. The missing standard in this story often turns out to be a background of beliefs, expectations, classifications, and attitudes about the items portrayed. That background will also include one's familiarity with other representations; art's "imitation" of nature rests on our construal of nature, and at least part of that is conditioned by art's own legacy of images.

I think that this simple observation has important consequences for our understanding of representation, but neither a differentiation of representation from imitation nor a symbol theory of depiction is among them. For all that has so far been said about representation applies equally well to imitation. Given an object and an order to replicate it closely we shall be as much limited by our beliefs and construals concerning it as by our skills and the availability of materials. For granting the impossibility of emulating in our copy all of the properties of the original, our actual ideas of a fair or close imitation are as much riddled with cognitive and pragmatic factors as that of a good representational likeness was said to be. Emulation of archetype in scale, weight, and strength would likely count toward close imitation of a fabric or an article of furniture; emulation with regard to such properties as color, texture, odor, age, etc., will be essential, important, or irrelevant to a close copy, depending upon the kind of thing copied and the use to which the replica is to be put. Thus the familiar observation that the artist intelligently *selects* among aspects and characteristics of his subject-matter, rather than "slavishly imitating" it, fails to distinguish either art from nonart or representation from imitation or copying. And although such artifacts may be products of, and evidence for, symbol-using activities of their makers, nothing so far indicates that the artifacts have themselves a symbol-like function.

II. "IMITATING APPEARANCES": SUPER-PICTURES

In praise of Masaccio, Vasari wrote: "He recognized that painting is naught else but the imitation of things as they are."[7] The second anti-imitationist principle denied that faithful pictures could be said to yield even imitations of things as they *appear* under prescribed conditions. In a sympathetic effort to understand the idea of a pictorial imitation of something in even *an* appearance, I shall introduce the idea of "super-depiction."

A super-picture of something would be, roughly speaking, a picture of it which, when viewed under some set of objective and subjective conditions, is indistinguishable to all from the thing itself, as viewed under some (possibly different) set of conditions. There would be many difficulties confronting efforts to refine this notion sufficiently for use, even in a relativized version: "P is for S a super-picture of X . . ." For instance, a super-picture of X must be super as an X-representation, so qualifications would have to be framed to exclude as super-depictional cases like that of a picture of a table which looks just like a table when placed face-side down on four table legs, or a portrait of a man which, when viewed in the dark, is indistinguishable from that man (as viewed in the dark). And of course qualifications on the "conditions of observation" would have to exclude subjective conditions (hypnosis, drugged states) on which any picture would be indistinguishable from its subject and all depictions would be super. I believe that the super-picture has served as an ideal for some theories of perspective and of picture-perception.[8] My question is whether this ideal—even if it were shown coherent by a precise specification — could explain our usual sense of terms like "realistic," "naturalistic," "convincing," etc., and whether its use might show that these notions are somehow related to the idea of imitating an appearance.

---

[6] These two points are distinguished by Nelson Goodman in his review of Ernst Gombrich's *Art and Illusion*, in *The Journal of Philosophy*, vol. 62 (1960), p. 59.

[7] On Wölfflin's translation of the passage in Vasari's *Vite* in *Classic Art*, tr. by P. and L. Murray (London, 1952), p. 4.

Pictures that we see every day in books, magazines, television screens, etc., are not super-pictures of what they represent. Pictures representing sunlit scenes are printed in low tones of black and white; objects are indicated in lines on what are obviously flat, limited, surfaces; closeups of human features fill a theater screen two hundred feet away. Yet we familiarly judge such representations to be more or less faithful or convincing, to be more or less lifelike, realistic, naturalistic. What could these judgments have to do with the imitation of an appearance? The answer might seem to be expressed in the following procedure: For each motif we imagine a super-picture, and then rank nonsuper-pictures of various styles, sorts, and conventions as they more or less closely approach this ideal of imitation. The super-picture perfectly reproduces an appearance, and ordinary pictures tend toward realism or lifelikeness as they approach super-depiction.

Now for different sets of conditions for observing the chosen subject there will be different sorts of super-pictures of it, and (as previously remarked) we would not in fact count all such pictures as equally faithful or lifelike depictions of the subject, since we would not count all such appearances as equal regarding it. We might then attempt to specify a single ideal by altering the description of the subject so as to incorporate part of the conditions of observation, and thereby to avoid the issue of a multiplicity of appearances: "Fuji — as seen through a window of this hut on a clear morning," "Henry — as he appears through a 35mm lense," etc.

The idea would again be that of ranking nonsuper-pictures for faithfulness by deciding which yield relatively closer imitations of an appearance, where "is a closer imitation" is defined as "more closely resembles the appropriate super-pictorial imitation." But of course for every chosen view there will be innumerable possible pictures of different sorts which resemble its selected super-picture in innumerably different ways. And the *respects* of resemblance which we select as most important will determine for us which kinds of pictures more closely match the ideal. Thus even within the restricted frame of "*an* appearance" of a thing there is unclarity about the notion of that appearance which is to be approximated pictorially.

Besides, the resemblance clause of the criterion was not supposed to apply to points of resemblance between nonsuper-picture and super-picture which the latter also bears to its indistinguishable natural archetype. Otherwise, interposition of the super-pictorial ideal would be unnecessary: we might as well speak forthrightly of points and degrees of resemblance between nonsuper-picture and subject-matter. Resemblance to the ideal picture, then, was rather to be in terms of those *pictorial* characteristics and devices which produced the super-depictive effect: emulation in size, perspective system, pigment distribution, for example. Yet such emulation of a super-picture would give no guarantee of faithfulness for what is depicted. In fact — as the checkered history of debates about perspective teaches us — it will often be necessary to depart from emulation of a super-picture in certain pictorial features to bring off just some effect in our ordinary picture which we admired in the super-picture. (Thus while linear perspective may win on optical grounds for a certain sort of super-picture, we may wish, in our easel-sized nonsuper-picture, to bend parallel lines with a "synthetic perspective" to recapture the panoramic effect of the large super-picture otherwise lost in the smaller format.)[9]

To sum up: In admitting the possibility of super-depiction I allow faithful representation by imitation of appearances. What the foregoing denies is that the notion of imitation as indistinguishability provides even the beginning of a *general* explication of what is involved in the capturing of appearances in pictures. Now a mimetic theorist may at this point object to the use of the limiting perfect case of. imitation: After all, to know what imitation leather is we need no notion of a perfect imitation leather. He may therefore still wish to explain statements of the form

---

[8] Cp. James Gibson's notion of "fidelity": "A faithful picture is a delimited surface processed in such a way that it reflects (or transmits) a sheaf of light-rays to a given point which is the same as would be the sheaf of rays from the original to that point," with "indistinguishability" as the limiting case ("A Theory of Pictorial Perception," *Audio-Visual Communications Review*, vol. 2 [1954], p. 14).

[9] For discussion see G. ten Doesschate, *Perspective* (Nieuwkoop, 1964); D. Gioseffi, "Optical Concepts," *Encyclopedia of World Art* (New York, London, Toronto, 1967), vol. 10, col. 766–769; J. White, *The Birth and Rebirth of Pictorial Space*, 2nd ed. (London, 1967). Maurice Pirenne's recent *Optics, Painting, and Photography* (London, 1970) — esp. pp. 93n1, 148f, 165–183 — supersedes his earlier writings on the subject of perspective. Pirenne's discussion there of "subsidiary awareness" of the picture surface is of special importance to this issue

> *P* is a perceptually faithful depiction of (an) *X*

by something of the form

> *P* depicts (an) *X* as possessing, and closely *imitates* (an) *X* in, perceptually significant aspects of its appearance,

which yields to our arguments about faithfulness while still getting imitation into the formula.

I believe that it is at this point that many writers deploy a principle like the second in Steinberg's account to remind us that aspects of appearance must be "translated" into pictorial terms and not copied. Unfortunately, their message is often obscured by carelessness about cases. Hermann von Helmholtz, for example, once remarked how because of physical limitations in his medium "the artist cannot transcribe Nature; he must translate her";[10] but many of Helmholtz's own examples of "translation" amount only to adjustments in the physical stimulus designed to reproduce the "subjective impressions"[11] we would be getting from nature: We color our picture-mountains blue since we cannot rely on the scattering of light to produce this effect as in nature, etc. Such "translation" is not an alternative to imitation. I believe that those who speak of representational "signs," "symbols," "structural equivalents in a medium," etc., usually have the rather different point in view, that a representational style or convention can itself act as a sort of medium or perceptual context upon which pictorial appearances are causally dependent.

We all know how physical context may affect the appearance of a color or a shape: How could style or system have a like influence? It would not suffice to say that style factors may be causally effective in conveying effects. A red figure on a green ground may look different from a red figure on a purple ground, and use of green with red may be a defining characteristic of your style. But although it could then be said that stylistic features determine the appearance of red in your pictures, knowledge of your style is not itself a perceptual context, since acquaintance with it is unnecessary for getting the optical effect in question.

To determine when style and knowledge of style acts as perceptual context we might observe the effect which misclassification as to style or medium has upon the appearance of a representation—and, thereby—upon its fidelity. Fred may closely resemble Harry by being like him in perceptually significant ways, and Fred's snapshot of Harry may also be said closely to resemble Harry, but the kinds of resemblance are different, for Fred resembles Harry as a man resembles a man, and Fred's picture of Harry resembles Harry as a (sort of) photograph resembles a man. The photograph might not be thought to resemble Harry even slightly if it were mounted on a man's neck, so as to be seen as a part of his body. We might expect then that the size, flatness, gloss, and color would so interfere with the overall resemblance that a poor photograph would come out no worse as a likeness. Such is an extreme case of the effects of misclassification upon pictorial appearance and resemblance; less bizarre cases have been described elsewhere in enlightening ways.[12]

A revision of imitationist accounts of good pictorial likeness in view of the preceding considerations may run to something like the following:

> *P* depicts (an) *X* as possessing, and either imitates *X* in or yields "symbols" or medium equivalents for, perceptually significant aspects of its appearance.

Possibly this is no less than what "the imitation of appearances" meant all along. Further assessment of the imitation model or metaphor would require then more careful analysis of the concept of imitation than its critics have so far afforded it. But I proceed now to a different question.

III. REALISM AS VIVIDNESS

What is the legacy of considerations like those preceding for our everyday notions of realistic, lifelike, naturalistic, or convincing depiction? Proponents of "language" theories of depiction have not agreed as to the answer. Having himself authored a forceful and influential version of that theory, Ernst Gombrich had occasion to complain in the second edition of *Art and Illusion* about a "group of readers [who] have sought support from this book for the . . . view, according to which the demand for fidelity to nature must always be meaningless, since everybody sees nature differently."[13] Now by his efforts as a historian (e.g. in *The Story of Art*[14] and *Norm and Form*)[15] Gombrich, like others, has made a large stake in what he calls the "objectivity" of notions of realism or fidelity of representation. Thus we find him in *Art and Illusion* attempting to explain the history of the conquest of appearances from style to style in Western art, and grappling unsuccessfully with formulae designed to protect these ideas from his own best insights about the conventionality of vision and depiction.

---

[10] "On the Relation of Optics to Painting," tr. by E. Atkinson, in *Popular Scientific Lectures*, M. Kline (ed.), (Dover publications edition, 1962), p. 285.

[11] *Ibid.*, p. 278.

[12] K. Walton, "Categories of Art," *The Philosophical Review*, vol. 79 (1970); R. Arnheim, *Art and Visual Perception* (Berkeley and Los Angeles, 1954), pp. 161–165; E. Gombrich, "Expression and Communication," in *Meditations on a Hobby Horse* (London, 1963).

Many would believe with Gombrich that there is more lifelikeness in a Vandyke portrait of Charles I than in an image of William in the Bayeux Tapestry, more truth to appearances in a photographic advertisement for a piece of electronic equipment than in a diagram for one, more naturalism in the Pergamon altar relief than in an early archaic kouros statue, more realism in a film sequence of a running horse than in a four-year-old's drawing of one, etc. Could such judgments tell of more than their authors' style prejudices? By a strategy which would further exploit rather than diminish the language metaphor for methods of representation, I shall argue that they can.

The relativist answer has had forceful expression. According to Paul Feyerabend, "realism of representation is an impossible doctrine" which "assumes that there is only one correct way of translating occurrences in the . . . world into situations portrayed in an altogether different medium."[16] In his recent *Languages of Art* Nelson Goodman holds that "realism is relative," that "'naturalism' is a matter of habit"[17] which reduces to the relative ease in interpretation of representational symbols, having earlier argued that

> Each way of painting represents a different way of seeing; each makes its own selection, its own emphasis. . . . What we regard as the most realistic pictures are merely pictures of the sort that most of us, unfortunately, are brought up on.[18]

Before sketching a case for an objective concept of pictorial realism or naturalism which I believe is adequate to the needs of an art historian such as Gombrich, I wish to make some preliminary remarks about cross-cultural arguments concerning ways of seeing. First, beginning with a familiar culture (our own), we shall find that neither ease at interpretation, absence of a sense of distortion, nor effect upon our nonpictorial seeing suffices for realism. We "read" cartoon drawings (like *Miss Peach* and *Peanuts*) with at least the ease — and through as much habituation — as we do drawings by Ingres or Signorelli, and without a sense of grotesqueness in distortion. Perhaps then realism is a (relative) property of what is in some sense a standard method of representation or reportage and a preferred way of perceiving, rather than simply an accustomed one.[19] But today's newspaper contains several sorts of representations, each familiar and standard to its own function. And although we may decline to select the one "best" of these, we might allow that they are not all equal in realism. What is wanted then is further explanation of what we mean here by "realism," "naturalism," etc. For the historian like Gombrich who would describe and explain characteristics, changes, and contrasts in styles by these terms, we need an account of them, free from appeals to suspect notions such as "literal likeness" or "imitation of appearance," but showing them fit for objective use.

Secondly, it seems an exaggeration to suppose that where the style is unfamiliar learning to "read" a picture will always effect one's seeing of nature. Sometimes a representational method resolutely avoids resemblance and interest in appearances. And even when there is such interest, learning to look at small or flat pictures, upside-down pictures, wide-angle photographs with marginal distortions, or at shiny daguerrotypes, need involve no change in one's nonpictorial seeing. It may involve, as earlier argued, an effort at seeing the new kind of image as an image of that type, rather than of some more familiar type, before accommodation breeds invisibility of certain stylistically invariant features which are counted irrelevant to resemblance. But whether increased fluency with the unfamiliar sort of image issues in some change in our way of perceiving will depend upon the particular case. Examples of misconstruing or stumbling about with an unfamiliar style may therefore not always be what the relativist supposes, and the mere fact that another culture would have rejected as distorted the pictures we choose as most faithful does not by itself show that its accustomed way of picturing reflects a distinct way of seeing. The analogy of accustoming one's ear to an unfamiliar accent, or of learning to read upside-down pages, may sometimes serve better than the model of learning a new language.

---

[13] *Art and Illusion*, 2nd (rev.) ed. (New York, 1961), p. xii.

[14] (London, 1950), esp. the treatment of Greek art, Renaissance art, and Cézanne.

[15] (London, 1966), esp. the title essay, with its notes.

[16] "Problems of Empiricism," in *Beyond the Edge of Certainty*, ed. by R. Colodny (Englewood Cliffs, 1965), p. 221.

[17] (Indianapolis and New York, 1968), pp. 36f, 231. The present essay, restricted as it is to the topic of perceptual fidelity in representation, avoids the wider issues involved in Goodman's general and impressive case for pictures as symbols.

[18] "The Way the World Is," *The Review of Metaphysics*, vol. 14 (1960), p. 52.

[19] See *Languages of Art*, p. 37f.

These remarks read the "seeing" in "ways of seeing" in a very literal way, and it may be urged that the realistic picture is not necessarily (but only for some cultures) the one which captures appearances: Rather, it is the one in which the subject seems most *real*. And for diverse theoretical reasons — if you are a Byzantine, a Moslem, or a Neoplasticist — appearances may be a poor key to your reality. I take this observation as limiting or describing, rather than as discrediting, the present task. Waiving the topic of other equally defensible uses of terms such as "realistic," "naturalistic," etc., the question is simply how well we can fare when we ask, with Gombrich, for the meaning of such expressions, interpreted in the sense of the rendering of perceptual likeness or of the capturing of appearances.

Gombrich's own reply is ambiguous between pictorial illusionism and fullness of information content, but neither alternative is acceptable.[20] The fact that a drawing by Eakins is more realistic than is a Greek geometric-figure vase does not mean that anyone is at all likely to be fooled by the Eakins. And from the fact that a picture represents a house as possessing many perceptual properties it by no means follows that the image looks much like a house. Yet I do not think that the notion of realism or naturalism in depiction reduces simply to that of good overall or physiognomic resemblance. The recent caution about the "real" in "realism" suggests to me that lack of distortion and good or significant likeness are no more sufficient for realism than is habituation. The fact that my picture is a poor likeness of Henry does not mean that it is an unrealistic picture, for we might still wish to call it a realistic picture of a man, or a naturalistic and convincing rendering of movement, space, or texture. And while we might not wish to characterize a surrealist landscape such as Dali's "Agnostic Symbol" simply by the label "realistic," surrealist images often derive their force from their realism, from the fact that their very distortions are made perceptually undeniable in their vividness.

Our criteria for realism, then, seem to favor sensorily vivid representation of a particular sort of characteristic, for any subject-matter. There is a difference between representing something as possessing a property and managing to convey, represent, or express those properties (such as brightness, hue, weight) or effects (movement, sparkle of light, etc.) themselves. In the latter case we speak of capturing, conveying, or rendering a quality or effect, or of giving a vivid sense of it. A like distinction obtains between resources of verbal descriptions — between vivid description and mere description-as. One question we can ask of a verbal description concerns the accuracy and inaccuracy of its attributions; another — parallel to the one about pictures — concerns the vividness of the description, whether it merely attributes characteristics or gives a vivid sense of them, so as to make the described situations seem real. Neither question concerns merely the ease with which we can read the description.

I shall call a pictorial device a *realism device* if it yields a vivid sensory equivalent, or gives a vivid sense of, a physical property or effect. Here I take into account both the manner and the object of representation or expression. Any such device gives us the sense of seeing, touching, feeling, hearing, etc., the objects, characteristics, or events represented. While whatever is conveyed in this way is thereby represented and conceived in a sensual manner, the thing conveyed need not itself be physical or perceivable: Consider, for example, the things expressed or represented with sensory impact by Gothic stained-glass or by a Rothko canvas. There is more involved in being a person than in being a physical existent, and characteristics which a person has as a person and not simply as a physical existent can be represented vividly. But for realism in representation we usually emphasize the expression of those characteristics which a person has as a physical existent: weight, texture, volume, movement, etc. The Western tradition in realism has favored rendering of these "tactile" characteristics as counting toward a sense of the material presence of the subject portrayed, but it would be possible to emphasize vivid conveying of auditory or chemical-sense characteristics as well.

---

[20] See *Art and Illusion*, pp. xiif, 99, 276, 299, and Goodman's criticisms in his review of Gombrich, p. 599.

There are many familiar realism devices as here defined. Motion pictures create a powerful sensation of motion by stroboscopic and other devices used on a succession of images; stereoscopes produce a sense of space through the use of two slightly differing images viewed by separate eyes at once, etc. Less powerfully, but not less interestingly, devices such as perspective distortion and diminution, overlap, shadow modelling, aerial perspective, etc., are used to create a sense of spatial depth. It is here rather than in its uniquely corresponding to "the way we actually see the world around us" or even in its being "the best possible approximation" to appearances — as some have claimed[21] — that a perspective system such as the classic "linear" method wins realism for a picture. With those who argue for the absolute advantage in realism of that perspective system over a frontal isometric system, or a non-perspectival system, we can say that perspective distortion and diminution are, with respect to realism, more than "conventional" devices. It would not then be necessary to introduce questions of the appearance of the nonpictorial world or of the indistinguishability of picture and motif to make this claim for artificial perspective. Nevertheless, it is to be expected that in their efforts at realistic effects, picture-makers will investigate nature to discover the cues for perception of space, weight, motion, etc., that they can exploit in their own creations. In this way artists discovered facts about optics and the psychology of perception which hold for nonpictorial perceiving.

A picture may then be described as a realistic or naturalistic picture of something in those respects or to that degree in which the picture possesses realism devices which convey strong sensations of those physical properties which it depicts the subject as possessing. And a pictorial style or convention is realistic to the extent that, or in those ways in which, it is characterized by such devices.

By supplementing this account with other criteria for realism,[22] we should be able to reconstruct the family of concepts of realism, naturalism, or convincingness which are familiarly used in discussions of the interest or lack of interest in the capturing of appearances in the history of art. To describe episodes of realism in art as marking renewals of interest in natural appearances would seem unavoidable, though to put them forward as periods of "the imitating of appearances" is a mistake. Consistent with a relativistic view of resemblance we may observe that there have been particular periods in the history of image-making where special interest was taken in ways of giving vivid impressions of sensible aspects of the world. Thus (it is often held) the Greeks discovered means for rendering space and motion, the Renaissance Italians discovered these effects and others (including Berenson's "tactile," "visceral," and "respirational" ones). And, on the present account, a Northern painter or a modern "magic realist" is called a realist not because he provides information about just what can be seen of an object from a given vantage point, but rather because he gives us so much detail and in such a way as to convey a keen sense of surface textures.

IV. Objectivity

Gombrich had wanted to secure a measure of neutrality and objectivity for descriptive terms such as "realistic." The vividness account of realism in depiction would allow objectivity in comparisons of the degrees and respects of naturalism or realism in children's drawings, Middle Kingdom Egyptian paintings, stick-figure drawings, blueprints, various sorts of photographs, etc., without having to argue for the literalist's "innocent eye" or privileged ways of perceiving. Such judgments get objective sense because, far from being bound by one's own habits or preferences, their responsible utterance presupposes acquaintance with the styles or conventions described. On this account we cannot tell about realism in a picture unless we can tell what sensory effects it would have. And we cannot tell this until we have grown sufficiently accustomed to a style that it can disclose its power to us, or until we have indirect evidence of what powers that style had for those who were fluent with it. Thus responsible comparisons for realism between representational systems presupposes either one's own fluency with them or evidence about the people who were fluent with them.

---

[21] See Pirenne's defense of linear perspective in "The Scientific Basis of Leonardo da Vinci's Theory of Perspective," *British Journal for the Philosophy of Science*, vol. 3 (1952), p. 183.

[22] Resemblance or lack of distortion is usually a factor. Besides this, existence of things of the sort depicted is sometimes a requirement, as in Courbet's remark, "Show me an angel and I will paint one."

This approach makes much of the authority of indigenous construals, while previous arguments here had urged tolerance regarding the diverse possible ways of perceiving nature in terms of representations of it. Why should we now speak of preferred ways of perceiving the *representations* themselves? The question is whether I am not overtaxing the linguistic metaphor for representation. Certainly, if we are to comment on the nuances of a sentence in an Ananda dialect we must either be ourselves fluent with that dialect or else rely upon the reports and behavior of those who are. But can we assume that representational systems are like descriptive ones in this respect?

The answer is that we can consistently speak of preferred ways of interpreting representations, so long as we remember that things are preferred for certain purposes, and that there are different possible purposes. The question, "How shall I look at this picture?" is a practical one, and it yields different answers depending upon our varying needs and interests. Suppose that my interest is simply in getting more enjoyment from a picture: The best counsel might then be for me to see the picture in whatever way it comes out best for me then or later. Toward this end I might go to a critic or a historian for guidance, to have pointed out to me what I enjoy perceiving but would not have noticed on my own, or to hear those anecdotes about the picture's history which make my present experience of it or my later thought about it more rewarding. No preference need be given in such cases to the original intention of the work. There are times when one may remark truthfully, "I liked it better before you explained it to me," and it is possible that a certain work will come off better for a given person, with certain interests and abilities, perceiving the work under certain conditions, if he does not attempt to approximate someone else's experiences. The fickleness of fashion regarding these abilities, interests, and conditions is attested to by the variety of responses over the years to incidents like the cleaning and brightening of favored works by "old masters."

But for the scholar who uses a society's legacy of images as evidence for judgments about that culture's beliefs, attitudes, perceptions, and doings, the situation is different. Correctness of interpretation becomes a meaningful and important idea here, and accordingly, insularity or prejudice of perception becomes a vice — largely because it leads to error. In these contexts, and with these purposes, the scholar's judgments about pictorial realism or naturalism are typically related to remarks about socio-economic, political, scientific, religious, and aesthetic matters. And in principle there is no more theoretical problem of objectivity concerning the question how convincingly something is conveyed in a picture than there is about the questions of what is represented there and what that thing is represented as being. I would be inclined here to appeal to facts about a picture's representational "conventions," but the fact is that there are no rules prescribing such conventions.[23] For conveying and expressing, as for representing-as, the "conventions" exist only in the greater or lesser agreement in the responses of those numbered among the fluent. Here the art historian finds, as for verbal languages and other cultural entities, communities with debatable and shifting borders which only approximate unanimities of response.

University of Western Ontario

Received May 26, 1971

---

[23] See E. Panofsky's discussion of the difficulty involved in defending the seemingly correct remark that van der Weyden's "Three Magi" represents the Infant as making an apparitional appearance: *Studies in Iconology* (Oxford, 1939), pp. 9f.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 3, July 1972

V. DELIBERATION AND THE REGULARITY OF BEHAVIOR

DIOGENES ALLEN

I WISH to examine the process of deliberation in order to specify precisely an issue in the free-will and determinism controversy. Let us begin by noting the significance of the question: In deliberation, is the agent engaged in the process of *setting* or *deciding* on the goal he is to pursue, or is he instead, given the antecedent circumstances and the person he is, *discovering* the goal he is to pursue? An examination of some arguments for each alternative indicates that the latter alternative is apparently the more plausible, and therefore that the libertarian who views deliberation as a process of setting goals is in serious difficulty (Sect. I).

I shall argue here that these arguments for deliberation as a process of discovery fail, and that the reasons for their failure greatly clarify the entire matter of the explanation of behavior. Underlying this contention is the assumption that the explanation of human behavior rests on a bedrock of behavior which people of a culture regard as "sensible" and "reasonable," or at least "understandable" behavior; behavior is thus reducible to regularities, regularities perhaps capable of systematization, but nonetheless these explanations or regularities only occasionally lend themselves to the making of predictions (and then, for the most part, only rough ones).

I. THE DIRECTION OF ATTENTION

Let us begin with a brief characterization of the process of deliberation. It involves, in part, a consideration of possible goals one might pursue, such as losing weight or learning to be a key-punch operator; a consideration of how to attain various goals; and what various courses of action require of one and their consequences. There is the attempt to balance all these factors so as to choose that goal and to follow that course of action which best satisfies what one wants. When one has determined or discovered the goal and the course of action most suited to that goal, one sometimes has to make an effort because, to follow a course of action that leads to a goal is hard or distasteful. In some cases one can and does make the effort. But in some instances, one comes to realize that the effort is more than one wants to make after all *vis à vis* the gain that is to be had, and one changes one's mind about the route or even the goal.

Sometimes one is not clear about what it is that one wants, and often becomes more clear about the relative value and hence order of one's wants, after reflection; nor is one always clear about what alternative courses of action are available and their consequences until more information is sought and brought into the process of deliberation. Without deliberation, then, the matching of wants and courses of action may be haphazard; to match them deliberately is to act rationally in the sense that wants and courses of actions are examined and an effort is made to match them coherently.

This characterization of deliberation has a very strong air of calculation about it, as though deliberation were a process of discovering what one's wants are, their relative priority, and which alternative means for their satisfaction is the most efficient. Even though an end has to emerge in the process of deliberation, our wants are counters which reveal their relative strengths and conflicts in deliberation, and are ordered by one's discovery, upon reflection, of what it is one really most wants. The whole process seems to be one of fitting together a number of items whose fit was not known prior to the deliberation.[1]

---

[1] Later we will consider the claim that, when by deliberation, one has reached the conclusion that a goal and a course of action is the best or most rational, given one's wants or desires, the agent must now choose whether to pursue the best or most rational course of action. Here the deliberation would be between doing or not doing what one has already determined or found to be optimum. This case is separated from the process of deliberation described in the main text, since a libertarian such as C. A. Campbell considers only this sort of case as vital to the libertarian and concedes the deliberation process to the determinist. Libertarians such as R. L. Franklin and Rem Edwards concede neither to the determinist.

Austin Farrer, Rem Edwards, and R. L. Franklin have objected to the interpretation of deliberation as a process of calculation.[2] Were it only a matter of calculation, they claim it would make deliberation a matter of *discovering*, not the *setting* of goals and courses of action. Edwards and Franklin would amend, therefore, the description of deliberation which I have just given to include the direction of attention in deliberation.[3] They claim that in deliberation one's very attention to a want and to a possible course of action and its consequence has the effect of elevating, lowering, ordering, and re-ordering one's desires; the very act of attention in deliberation affects all the items being considered by affecting the strength of one's desires. Thus deliberation is not merely a process of discovering by calculation which of the available courses of action would best satisfy one's given wants, but sometimes it is also a process whereby one affects what it is that one wants and it is thereby a process of *setting* goals.

The point of this objection is that were deliberation merely a matter of calculation, one's course of action would have been determined by antecedent factors prior to the act of deliberation. Given all the external circumstances and given the person the agent is, deliberation would only be a process whereby the agent finds out which course of action his own wants and the situation dictate. Deliberation would still be significant in that by means of it one would much more likely act reasonably and coherently, but it would not be a process of setting a goal.[4] Edwards and Franklin think that deciding what one is to do is determined by deliberation and not by factors antecedent to it.

There are two objections which may be raised, however, to show that the modification of the description of what takes place in deliberation proposed by Edwards and Franklin does not materially affect the situation. That is, it does not require modification of the view that the course of action is determined by the external conditions and the person as he is prior to deliberation. One of them can be presented by drawing upon an article by John Hospers.[5]

Hospers argues that some people (neurotics), given their heredity and environment, perform particular deeds no matter what they try to do and no matter what rational considerations are brought to their attention by themselves or others. He also claims that even the so-called normal person is not to be given credit for his actions.

> Granted that neurotics are not responsible for their behavior (that part of it which we call neurotic) because it stems from undigested infantile conflicts that they had no part in bringing about, and that are external to them just as surely as if their behavior had been forced on them by a malevolent deity . . .; but the so-called normal person is equally the product of causes in which his volition took no part. And if unlike the neurotic's, his behavior is changeable by rational considerations, and if he has the will power to overcome the effects of an unfortunate early environment, this again is no credit to him; he is just lucky. If energy is available to him in a form in which it can be mobilized for constructive purposes, this is no credit to him, for this too is part of his psychic legacy. (P. 37.)

He then considers the objection that some people can change their unfortunate behavior from the direction toward which their heredity and environment dispose them to act, and that this is not a matter of luck.

> But, one persists, it isn't a matter simply of luck; it *is* a matter of effort. Very well then, it's a matter of effort; without exerting the effort you may not overcome the deficiency. But whether or not you are the kind of person who has it in him to exert effort is a matter of luck. (P. 39.)

To have the ability to have one's behavior changed by rational considerations brought to one's attention by others as well as by oneself, and to have the ability to exert effort are the result of previous conditions over which one now has no control and, presumably for Hospers, over which one never had any control.

---

[2] Austin Farrer, *The Freedom of the Will* (London, 1958), e.g., p. 39. Rem B. Edwards, "Is Choice Determined by the 'Strongest Motive'?", *American Philosophical Quarterly*, vol. 4 (1967), pp. 72–78. R. L. Franklin, *Free Will and Determinism* (London, 1968).

[3] For Franklin, it is the core of his version of libertarianism.

[4] What I am describing is sometimes called the course of action which is the most economical vector; that is, in relationship to all the internal and external stimuli, it is that behavior which will be most deeply satisfying. A free person has often been defined in terms of one who follows (or approximates) the most economical vector, in contrast to a person who is defensive, and denies or distorts some of the relevant data, and whose course of action is thus less than fully satisfying. Both courses of action are said to be determined by what preceded them, but only the former course of action, because one gets what one wants (defined in terms of degree of satisfaction) is free. See, for example, Carl R. Rogers, *On Becoming a Person* (Boston, 1961), pp. 192–193. This is not the sense of freedom supported by Farrer, Edwards, and Franklin.

[5] "What Means This Freedom," reprinted in *Free Will and Determinism*, ed. by Bernard Berofsky (New York, 1966).

This suggests that the point made by Edwards and Franklin about the effects of attention on desires and goals in the process of deliberation, and hence on the course of action seen as the most suitable, does not materially affect the description of deliberation as a process of discovery. For Hospers grants that the course of action selected is indeed affected in those cases in which the agent has the power to deliberate and to make an effort, but these capacities of the agent are not the result of matters over which he has control. His heredity and environment have made him what he is; they have given or withheld certain capacities. So even if he has the power to deliberate and to make an effort to carry out the upshot of the deliberation, still he is not to be given credit for what he does. The agent's capacities make a difference to what is done, but the agent is not responsible for having these capacities. So given the agent you have got, given all the other prior factors and the present situation, the course of action is a result of a calculation which matches wants with what is available, and deliberation is to be considered a process of discovery.

There is, however, an ambiguity in Hospers' position which prevents its employment as an argument to downgrade deliberation from a process of setting goals to one of discovering them. One could respond to Hospers' position: regardless of how one gets a capacity, the exercise of it is dependent on the agent. He may use it or not use it, and the extent of its use is at least partly up to him. This response is possible because Hospers has not made it clear whether it is the existence of a capacity only, or its use and degree of use which are also dependent on heredity and environment. He grants that some people are "capable of exerting effort," have the "power to overcome the effects of an unfortunate early environment," and that some people's "behavior is changeable by rational considerations." But these expressions are consistent with the claim that the ability of using or not using these capacities (and their degree of use) is at least partly up to the agent. And they are also consistent with the view that the use or non-use of these capacities (and their degree of use) is determined by the agent's heredity and environment so that the use or non-use of a capacity is not to his credit.

To overcome this ambiguity one must turn to the second objection which could be made to the modification of Edwards and Franklin to the sketch given of the process of deliberation. The second objection is based on the view that every act is done for a reason, and for every reason which an agent considers for an action there is a reason why it is a reason for him and how significant a reason. We may ask then, 'Why did he switch his attention to *Y*?" A reason can be given for this, and then we may ask why that reason had the weight it did for the agent. The regress can be continued indefinitely.

The direction of attention was introduced because the outcome of the process of deliberation seemed fixed or determined by factors which had occurred prior to the deliberation; given the kind of person one is, one deliberates as one does, weighs pros and cons for an action as one does, so that the entire process is that of matching wants to a course of action that best satisfies them. Should one regard this process determined because there are reasons why every reason for and against a course of action is weighed as it is by a given agent, then to introduce the idea of our ability to direct our attention does not materially affect the account of the process of deliberation. One switches one's attention for a reason: because, for example, given the kind of person one is, one turns one's attention to regard A-type of considerations very closely, and ignores B-type reasons; or one keeps to the task of deliberation despite the ringing of the telephone, or one is distracted by a distant noise. This act of directing one's attention is as much or as little determined as other parts of the process which are done for a reason. Thus if one cannot meet the problem that because everything has a reason for its occurrence, and there are reasons why this reason is a reason, it will not help when one considers the process of deliberation to introduce the notion of the direction of attention; for this act is just as vulnerable to determinism as the other aspects of the deliberation process. We must then examine more closely what happens in the process of deliberation from the standpoint of the relationship between a reason for an action and an action, and why a reason has the weight it does.

II. The Ineffectiveness of a Regress Per Se

Consider a person who deliberates about giving up smoking. His wife wishes he would give it up because its odor is offensive and it lingers in the house and on his clothes. He often finds that it irritates his sinuses and gives him very bad headaches. His doctor has informed him that he is running a serious risk of developing cancer unless he stops smoking altogether. Yet he basically enjoys smoking and it helps him find something to do during long boring sales meetings.

Now let us say that over a period of time he considers and reconsiders the above pros and cons of continuing to smoke and of ceasing to smoke, and finally he decides that he is going to stop smoking despite his enjoyment of it and its occasional usefulness, primarily because of the cancer warning, but partly too because of the sinus headaches and as a bonus his wife will be pleased.

We clearly see that these are reasons for giving up smoking; if someone explained why he gave up smoking for these reasons, we would believe that his action had been sufficiently accounted for. But one can still ask, "Why are these reasons reasons *for him*, and why does *he* give them the weight he does?" If someone presses this question it is not because one does not understand the man's behavior or considers the reasons insufficient to account for or to explain the man's behavior. As we shall see in Sect. III, this fact is very important. Nonetheless when it is asked in the dispute over whether or not all behavior is determined, it looks as though one and only one course of behavior is truly open to a person. Whatever he does, there is a reason why he did what he did, and a reason why that was a reason for him, and so on, which is the same sort of regress as we saw with the direction of attention.[6]

There is no need to panic, however, at the idea that there are reasons why some things are reasons for the agent and reasons for the degree of weight they have for him. This regress does not *per se* make the situation any more or less deterministic. Presumably the reasons given for ceasing to smoke do not *appear* to determine the behavior, since if they clearly did, there would be no need for a determinist to begin a regress. We would already know that the action was determined by the given reasons. Moreover, if a deterministic relationship is not apparent between the action and, for example, a cancer warning, headaches, and a wife's objections, then one is not going to find that there is such a relationship *merely* by reference to the fact that there must be reasons why these are reasons for the agent and together of decisive weight. It depends on the *type* of antecedent reasons to be uncovered.

The regress may uncover reasons which are of a type that we know determine behavior. They would be causes which entail that a doctor's warning, headaches, and a wife's objections are reasons for him and together greater than any that run counter to them (enjoyment; something to do during sales meetings). But it may uncover reasons which are of the type with which we began (doctor's warnings, headaches, wife's objections, enjoyment, something to do at sales meetings), ones which do not appear to determine behavior. They would be reasons for giving greater weight to doctor's warnings, headaches, and a wife's objections taken together than to any that run counter to them (enjoyment; something to do during sales meetings), but they would not entail that such weight be given to them. If this were the case, then we do not have determinism now, any more than we did before introducing the idea of a regress. We simply find the same type of reasons in the regress that we had before we began it.

To maintain determinism, one must maintain that the antecedents to the reasons given by the agent,[7] which are to be found by a regress, will always be causes which determine behavior, or like causes and determine the outcome. But were it the case that we knew or could find a way to show that all reasons given for actions (in the case considered, giving up smoking) were like causes which determine occurrences, then we would not need to consider their antecedents. We would already know that the behavior was determined. A regress is introduced as an argument by a determinist precisely because we do not know whether these reasons are really causes or like causes which determine behavior; and precisely because we do not know, the intention of introducing the regress is defeated. We need not, then, be troubled by a regress *as such*.

III. Logically Complete Explanations

It is extremely difficult to decide whether or not the reasons used to explain behavior determine behavior. Many distinctions between types of reasons, between types of causes, and between reasons and causes have been drawn in recent years. But we still do not know whether these differences are the relevant ones for showing that reasons do or do not determine behavior. For it is possible to state the controversy between free will and determinism in such a way as apparently to nullify the significance of the various distinctions between reasons and causes as far as this controversy is concerned.

---

[6] This question arises at least three times in Franklin, who seeks to defend a libertarian position. He admits its impressive force against libertarianism, and each time he puts off engaging it until the last chapter of his book. We will consider his response in Sect. III.

[7] Assuming here that these indeed were the reasons he did give up smoking.

We can see this by considering Franklin's statement of the issue between libertarianism and determinism.[8] Given all the environmental circumstances, and the state of the agent (*which includes his reasons, intentions, conscious and unconscious motives, and the physiology of his body*), then what the agent will do is determined. The environment and the state of the agent (which together I will refer to as *X*) are a sufficient condition for *Y* (what the agent will do). A libertarian, on the other hand would say that given *X*, the agent might do *Y* or something else; that is, he denies that there is a nomological connection between the initial conditions *X* and what the agent will do *Y*. As Franklin puts it, the determinist asserts a regularity such that "... there is always some true proposition, even if unknown, which asserts or entails that, whenever this *X*, then this *Y*" (p. 14).

Were one to grant that there is always a reason why an action occurs, then it seems that the explanation must be such as to explain why *Y* happened rather than anything else. If *X* leaves open the possibility that *Y* or something else might occur, it does not explain why *Y* occurred. To explain why something happens, the explanation must be logically complete, that is there is to be no room for the question: "Granted all you have said, why did *Y* rather than something else occur?" Were there a nomological connection between the initial conditions and what the agent will do, as the determinist maintains, then we would have a logically complete explanation.

Franklin argues that there is more than one type of logically complete explanation. The type we have just mentioned in which, given the initial conditions, we can with nomological laws deduce what the agent will do, he calls theoretical explanations. But he claims there are also practical explanations in which the agent explains why he did something. In many instances, the reasons he offers are good ones; that is they would be good ones for anyone who was in a situation like his own, and would justify the action. Such reasons are said to be universalizable. That is, if anything is to be a good reason at all, even to the agent, then in a case where the agent and the situation are similar to the present one in all relevant respects, a similar decision must be the reasonable one again. With such reasons for an action, it does not make sense to ask, "Granted all you have said, why did you do *Y* rather than something else?" The question has already been answered.

The claim, then, is that the relationship of reasons to an action are not causal in the sense that given the reasons, and other relevant data, the course of action is deducible. The action is explicable because the agent's reasons are connected with what we all recognize as reasons for an action, given the situation.[9]

Franklin is nonetheless troubled by the determinist's claim, which he grants, that there is a reason for every action. When he considers the relationship of the result of the process of deliberation (the course of action which, given all the reasons considered, seems to be the best or most reasonable) and what the agent does, he wishes to maintain the freedom of the agent to act differently from the way his deliberation directs him to behave.

> A typical conclusion [to deliberation] would be that *X* is the best, or sensible, or appropriate thing to do. But from this what follows about whether the agent will in fact do *X*? Even the special case of *A*'s deciding that he ought to do *X* does not entail his deciding to do *X*. All that practical reason shows is that if he does *X* he can justify his conduct, and if he does not he cannot. Even to take the extreme case, if only one course of action attracts him at all, there is still no entailment that he will do it; men sometimes refrain from doing things though they can offer no reason at all. (P. 255.)

The result of one's deliberation, then, still leaves one with a choice to make. One has a reason for a course of action, but one may or may not choose to act on that reason.

---

[8] *Op. cit.*, esp. pp. 10–20, 244–259.

[9] We may be mistaken sometimes in regarding some reasons as giving an explanation of an action and as together forming a logically complete explanation. But such mistakes *in instances* do not rule out this *type* of explanation. The discovery of mistakes in instances would have to take the form of *empirical* discoveries of nomological relationships between initial conditions and behavior which replace a particular explanation of a practical kind. But to rule them out as a type, one would have to rely on arguments. In what follows, we do the latter. We consider the *argument* to the effect that the only full explanation of an action is one in which the explanation accounts for why it and no other act was done, and that this rules out practical explanations as logically complete explanations.

> There is indeed one sense in which we will always have a reason for what we do, namely that we can give a logically complete one. But in a more important sense in such cases we are deciding not what reason to follow but whether to follow reason. (P. 257.)

But Franklin must face the difficulty that there is an explanation for a decision to follow reason or not to follow it. On the one hand, if the agent chooses to do the reasonable act, *X*, were we to ask him why, he can cite all the reasons that went into his deliberation which show that the act is reasonable. He can thus produce the reasons he did *X*, and together they would be a logically complete explanation; for it would make no sense to ask: "Granted those reasons, why did you do *X*?" We have already given the reasons for doing it. On the other hand, let us assume that the agent chose not to do the reasonable act *X*. This too will have an explanation. It may be, for example, weakness of will, that is, when the time came to do *X*, other desires welled-up and re-shifted the previous balance of considerations which the deliberation had established. So there is also a reason here for what is done, a logically complete one. It is an unreasonable, and hence unjustifiable act, for it is done without a good reason; but nonetheless there is a reason for it. So whichever act is done, there is a logically complete explanation for it.

But we have not explained why the agent did *X* instead of *Y*, or *Y* instead of *X*, by citing logically complete reasons. We must give the reason why the logically complete reason on which he chose to act weighed with him as it did. But this has the effect of determining the act. To grant then that there is a reason for every act, as Franklin does, allows a regress to arise which prevents a logically complete explanation of the practical kind from fully explaining why an action is done without also determining the action.

Franklin admits that this regress creates a difficulty for him (pp. 233–235, p. 257, and pp. 259–260). He seeks to meet it by a distinction between constructive and critical metaphysics, and an argument from analogy. The constructive metaphysician uses the principle of sufficient reason to argue for the existence of a necessary being on the basis of contingent beings, and the critical metaphysical call for a stop to the regress before it reaches such a being. So too the determinist searches for the reasons why the reasons are reasons for a particular agent and why they have the weight they do for him, and hence form the basis for his choice of *X* rather than *Y*. The result to such a regress can only be reasons or causes which entail the course of action taken. The libertarian is like the critical metaphysician who says the regress must come to a halt before one and only one course of action is determined (pp. 285–301).

To seek to stop the regress seems to me to be unnecessary. For it can be shown that regularities, i.e., specification of one course of action rather than another by reasons, do not entail determinism. We can deal with the difficulty by making a distinction between a situation before and after an action is done. In the above example in which we first took it that the agent did *X*, and then considered that the agent did *Y*, we were examining the situation *after* the choice is made and the action is done. In this case, when *X* was chosen and done, we had a logically complete reason for the act. When *Y* was chosen and done, we had another logically complete reason. But *before* the choice is made and any act is done, we cannot explain which way the agent *will* go. The reasons for the reasonable act and the state of the agent (that he is weak-willed) do not tell us what he will do. For though he may do *Y* because he is weak-willed, he may, even though he is weak-willed, do *X* (act reasonably). The factor (weak-willed) cited after *Y* was done to give a logically complete explanation of his unreasonableness, when considered before an act is done, does not determine what the act will be. For weakness of will is the type of reason which does not make it evident which action will be done. It is the type of reason which, though it can be used to give a logically complete explanation after an act is done, beforehand, it does not evidently specify which of the two acts will be done. Such reasons which allow us to give a logically complete explanation for an act *ex post facto*, then do not *ante facto* specify which act will be done. This is true also for the reason that would explain why weakness of will was overcome or not overcome after the act is done; when viewed prior to whatever is done it may be the type of reason which does not specify that weakness of will, will or will not be overcome.

Though the type of reasons now used to make up logically complete explanations of a practical kind do not specify which of the two acts will be done, it is possible that one day any one or all of these reasons will be found in fact to specify which of two possible acts will occur. That is, practical explanations will be replaced by nomological relationships between initial conditions and behavior. The reasons now used in logically complete explanations of a practical kind do not now appear to specify what will be done.

We can by this distinction then satisfy the principle: for every act there must be an explanation why it was done rather than other possible acts. Once done, we can in principle give a logically complete reason for it. Prior to its being done, we often may not be able to designate what will be done. Hence, reasons, *ex post facto*, do determine acts by explaining why one was done and not another; they do not prior to the act.

This view relies on the notion that there are indeed regularities in human behavior. Whatever is done by a person is in principle capable of being explained. Whichever way an agent chooses to act when he concludes his deliberation, there will be a logically complete explanation of why he did what he did. But the explanation will be *ex post facto*.

Examples of these regularities can be cited. In our culture and often to varying degrees in other cultures, the probable elimination of headaches, a doctor's warning about lung cancer, and the wishes of a wife are generally accepted as good reasons for stopping to smoke. Such reasons are universalizable, that is, any agent from our culture faced with these reasons and the counter-reasons we considered earlier, would regard the decision to give up smoking as a reasonable one (though these reasons may have more or less weight for different agents). They form a logically complete explanation for the act of stopping smoking; were one to stop smoking for these reasons the behavior would be fully explicable. But were the action not taken, this too is open to a logically complete explanation: people often do not follow the best or the most rational course of action because, for example, too many counter-reasons arise when the moment to act reasonably actually comes and the agent acts on those counter-reasons. But were those counter-reasons regarded prior to the act, we would not necessarily know whether they would be effective any more than we did in the case of weakness of will. So all behavior is in principle open to logically complete explanations and thus reducible to regularities. Within a culture or for a single personality type they may be capable of systematization. But this does not mean that on the basis of initial conditions, and logically complete explanations, we can say that action *X* will be done and not something else. These regularities only allow one to explain why *X* was done instead of something else *ex post facto*. All occurrences and actions are thus in principle explicable; but these regularities only allow rough predictions of behavior on the basis of what people often do.

IV. CONCLUSION

Regularities of behavior and full explanations of actions are consistent with the view of deliberation as a process of deciding what to do. The agent must in effect select between reasons for different actions, reasons which *ex post facto* form various possible logically complete explanations which will fully explain what was done. These logically complete explanations rest on a bedrock of reasons which people of a culture recognize as reasons, and when used to explain what is done, are recognized as a full explanation. That a full account of *whatever is done* is in principle possible does not in itself entail that, given the situation and the history and present state of the agent, he *will* do *X* instead of anything else. Regularities then do not force one into thinking of deliberation as a process whereby one considers reasons which determine one and only one course of action. So the existence of regularities do not form a reason to say that in deliberation the agent is merely discovering what course of action he will pursue, instead of deciding upon it.

Princeton Theological Seminary

Received April 2, 1971

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 3, July 1972

VI. AESTHETIC VALUE

JOEL J. KUPPERMAN

IN the closing chapter of *Principia Ethica*, G. E. Moore gives the following account of beauty:

> It appears probable that the beautiful should be *defined* as that of which the admiring contemplation is good in itself. That is to say: To assert that a thing is beautiful is to assert that the cognition of it is an essential element in one of the intrinsically valuable wholes we have been discussing....[1]

Moore thus analyzes "beautiful" in terms of "good": in effect he claims to give a formula that reduces aesthetic value to a function of ethical value.

In what follows I shall work upon Moore's formula. That is, I shall criticize the formula, suggest modifications, and in general attempt to extract the kernel of truth that it contains about the relation between ethics and aesthetics. I shall discuss the formula *vis-à-vis* only works of art; there are special difficulties with regard to natural objects which will not be discussed.

We may begin by discarding the term "beautiful." It may be that when Moore wrote "beautiful" was the most general term of aesthetic praise. Even today it may be possible to discern some works of art whose admirers are strongly inclined to call them beautiful: Botticelli paintings perhaps, or Vivaldi concerti or Wordsworth poems. The 20th century though has developed a taste for art that normally would not be called beautiful, and a critical vocabulary to correspond to this taste. One has only to think of the productions of such men as Stravinsky and Picasso, or Genet and Beckett, as well as the paintings of Bosch, or the plays of Webster and Tourneur. We praise these men in a variety of ways; but at the core is the contention that their art is "good," and has "aesthetic value." To inquire about beauty is, in the context of current usage, at best to inquire about one species of aesthetic excellence (and at worst may suggest damning with outmoded praise); a more general inquiry will concern itself with aesthetic value in general. Moore himself undoubtedly meant his definition to do justice to aesthetic value in general. Consequently I shall treat it as a theory about aesthetic value, and tinker with it accordingly.

Now if the aesthetically good is said to be that of which the admiring contemplation is good in itself, we confront certain difficulties. The following four seem to me to be most serious.

(1) There is an implicit paradox in the notion that aesthetic value can be defined in terms of the value of admiring contemplation. The objects that one ought to admire are those of high aesthetic value: that is (according to the theory), those the admiring contemplation of which are of high value. An aesthete, in determining which of a number of works of art are very good, would have to compare the values of his experiences of admiringly contemplating the works. Thus he would not know what to admire until he admired it. This suggests that if the aesthetic value of a work is defined in terms of the value of an experience of contemplating it, admiration should not be specified as an element of the experience. Admiration may be regarded as supervenient: i.e., as an element accompanying the aesthetic experience, and depending on its nature.

Do we then simply define the aesthetic value of a work in terms of the value of an experience of it? There is a very strong argument against this. Clearly not every encounter that someone has with a work of art can be taken as providing an indication of the value of the work. Someone may listen to a piece of music and not be able to specify those elements in the work which better prepared listeners recognize as the melody. Teachers of literature often speak of pupils as failing to "understand" poems or novels. If a schoolboy "gets nothing out of" a Picasso painting, this generally would not be considered to count against the Picasso. Only what one might call a "proper" experience of a work (one that involves "understanding") would normally be held to provide an index to the value of the work. Thus it is plausible to revise Moore's formula for aesthetic value to read as follows. To say that something has high aesthetic value is to say that the proper experience of it has high value.

---

[1] *Principia Ethica* (Cambridge, 1954), p. 201, italics Moore's.

(2) However, there is a general difficulty in specifying what counts as a proper experience of a work. Moreover, it is difficult to claim, even in theory, that there is only a single proper experience of any given work. Two specific problems here concern emotional response, and what are called "associations" of a work.[2]

It is clear that someone who feels absolutely no pity and fear while watching a Greek tragedy is failing to have a "proper" experience of the tragedy. The failure may be as serious as a failure to understand the literal meanings of some passages. We would be as loath to count against *Oedipus Rex* the low value of the experience had of it by someone who had absolutely no emotional response as we would be to count against it the low value of the experience had by someone who could not fully understand the relation between Oedipus and Jocasta. But what degree of pity and fear do we require for a proper experience of the work? Further, do we fault someone who fails to have the emotional response to a sentimental melodrama that the work clearly calls for?

The problem of association is even more thorny. It is generally agreed that there are what might be called "legitimate associations" to works of art, which enhance the value of the experience they provide. For example, a man may visualize some works of Byzantine art while reading Yeats's poem *Byzantium*. He may recall Hogarth's *Rake's Progress* while watching the Stravinsky opera of the same title. Critics normally consider themselves entitled to refer to such associations in their discussion of a work. The man who has a rich store of such associations may have a richer, and more valuable, experience of a work than the man who lacks this store.

The word "association" has the connotation of something optional. If something is vital to the understanding, or the proper appreciation, of a work, we do not refer to it as an association. We speak of it as part of the meaning of the work. For example: a critic may consider familiarity with the related passage in *Hamlet* to be essential to understanding the "Good night, ladies, good night, sweet ladies, good night, good night" passage in Eliot's *Wasteland*. He would then not speak of *Hamlet* as an association, but would simply speak of the related passage as providing part of the meaning of the Eliot text.

This creates a problem for Moore, or for anyone tinkering with Moore's formula. The problem is this. The man who visualizes Hogarth while watching Stravinsky may have a richer and somewhat more valuable experience than the man who does not visualize Hogarth while watching Stravinsky, but we do not thereby say that the latter failed to have a proper experience of the work. This suggests that there is a range of values, rather than a single value, which the proper experience of a given work may have.

If this range of values is wide, this opens the door to aesthetic relativism: it then becomes especially tempting to say "Work of art *A* is better than *B* for *X*, but for *Y* work of art *B* is better." In the dissertation already cited I argued that the range of values adhering to the proper experience of a given work could not be wide, on the ground that anything which made a great difference to the value of an experience of a work of art would be considered to be required for a proper experience of the work, rather than being classified as an association. I am no longer convinced that a formula somewhat like Moore's can be saved by means of this argument.

(3) A further difficulty is this. It is not altogether common that a man who is looking at a painting, listening to a piece of music, or reading a poem is so thoroughly preoccupied with the work that it entirely absorbs his consciousness. There are degrees of attentiveness and inattention, but in general we do not deny that someone may have stray thoughts or moods in conjunction with what genuinely counts as a proper experience of a work of art.

It follows that the states of mind of someone experiencing a work of art very often are a composite which includes aesthetic experience, but also includes other elements. One such extra-aesthetic element may be the admiration or contempt supervening the aesthetic reception of the work, but there may be other elements. Moore recognized that there may be other psychic elements accompanying aesthetic experience, but in my opinion he failed to deal adequately with the difficulty this creates. What if the extra-aesthetic elements contribute heavily to, or subtract from, the value of the total experience that includes aesthetic experience? One would not like to count that for or against the work of art that is being experienced. It may be that the experience of hearing a Beethoven quartet while having a tooth extracted is in some cases less valuable than that of listening to a Johann Strauss waltz just after having one's marriage proposal accepted. But this surely does not imply that the Strauss waltz has greater aesthetic merit than the Beethoven quartet has. At the very least, any formula like Moore's, which makes the value of the experience of a work an index to its aesthetic value, must stipulate that the extra-aesthetic elements of any experience which is taken as an index do not contribute very much to, or detract very much from, the overall value of the experience.

---

[2] These and related problems are discussed in great detail in my "Evaluations of Works of Art," dissertation accepted for the Ph.D. degree, Cambridge University Library, 1962. The major difference between this dissertation and the present essay is that in 1962 I took seriously the notion that *some* formula connecting aesthetic and ethical value had to be found, even if the formula Moore proposed was not acceptable. Thus the last half of this essay goes beyond the dissertation.

(4) It also may be difficult to draw a sharp line between the time when an aesthetic experience has taken place, and the time when it no longer is occurring. The aesthetician Stephen C. Pepper developed the concept of "funding" as an element in aesthetic experience.[3] He is clearly right in suggesting that elements of aesthetic experience may be cumulative and delayed.

Not only does funding undeniably take place in our continued experiences of a work, but it also is true that a work of art may, as it were, reverberate in our later experience of other things. Indeed a large part of the value of some works for us is the savor of them in recollection, as when we are confronted with events in life that remind us of events in a novel. Works of art may change our perceptions: witness the way landscape paintings may get us to perceive scenes in nature differently. There are a variety of ways, as I. A. Richards argued at length, in which good works of art may make our attitudes more sophisticated, and bad works of art may make them more crude.[4]

It is difficult to argue that all of this is irrelevant to the value of a work of art. But what then of a formula like Moore's? If we place the obvious, common-sensical time limit around what we call the experience of a work of art, we are excluding a good deal that is relevant to the value of the work. But if we loosen these limits, we run the risk of defining the value of a work in terms of the value of the entire life of the person who appreciates it; and this is so loose as to be useless.

The fourth difficulty especially convinces me that no formula like Moore's can be used to reduce aesthetic values to a function of ethical value. Even without this difficulty, one might be suspicious of the vagueness of terms like "proper experience," terms which would have to appear in any nearly adequate formula. Finally one may ask, "Why must we arrive at a formula?" Must philosophy mimic science by reducing aesthetic value to a species of ethical value, in the way in which a scientist might reduce, say, lightning to a kind of electricity?

Nevertheless there seems to be something in what Moore said. In the remainder of this essay, I shall point out two respects in which aesthetic and ethical value appear to be connected.

Before I begin I should say something about the conception of ethical value on which I rely. It is, for purposes relevant to this essay, the same as Moore's. Any statement that an experience is valuable, or lacking in value, or more or less valuable than another experience, counts as a claim of intrinsic value and as an ethical claim, no matter how many conditions the claim is hedged about with.

There are many varieties of usage of terms such as "ethical." and many philosophers hold a view of the domain of the "ethical" which is significantly different from that of this essay. Some philosophers simply equate "ethical" with "moral," or use "ethical" as an adjective for theorizing about the moral. But in the most common use of "moral," moral judgments concern duties and obligations, or worth that is intimately related to the performance of duties and obligations. If the "ethical" and the "moral" are coextensive, then neither the first book of the *Nicomachean Ethics* nor the last chapter of *Principia Ethica* could be regarded as part of ethics.[5]

Other philosophers would say that ethical judgments are distinguished from other kinds of evaluative judgments by virtue of the kinds of reasons that are advanced for them. A philosopher holding this view would reject *a priori* the argument of the remainder of this essay. I should point out

[3] *Aesthetic Quality* (New York, 1938), pp. 43–44.

[4] Cf. *Principles of Literary Criticism* (London, 1959).

[5] Among recent attempts to distinguish between "ethical" and "moral" I should mention P. F. Strawson's "Social Morality and Individual Ideals," *Philosophy*, vol. 36 (1961), pp. 1–17 and my own effort in *Ethical Knowledge* (London and New York, 1970), pp. 100–102.

that one may argue for the value of experiences in a variety of ways, that there is no single form such argument must take.

I owe the reader some brief account of how, in my view, ethical judgments can be distinguished from other kinds of evaluative judgments. (What follows is distilled from a very long account contained in work in progress.) Ethical judgments are distinguished by having what might be called a "conclusive" relation to conduct. If we are deciding what to do, the full range of correct ethical judgments that can be made about our situation will give us all we need for deciding what we ought to do, what is permissible, and what we ought to avoid. If we postulate a set of statements that has conclusive prescriptivity (a set consisting of claims concerning obligation and statements of values that can be gained or lost) any statement that is required to complete the set is an ethical statement.

Aesthetic statements of course do not have conclusive prescriptivity. Even the entire aggregate of aesthetic judgments that can be brought to bear upon a situation will not conclusively tell us what we should do. But this does not imply that there are no logical or epistemic connections between aesthetic and ethical statements.

(1) The logical connection between statements of aesthetic value and statements of ethical value may be illustrated by the following. Suppose that Jones tells Smith, "Play *X* is better than Play *Y*, and your time would be better spent if you went to Play *Y*." An explanation is called for. If Jones tells Smith, "Play *X* is better than Play *Y*, and your time would be better spent if you went to Play *X*," no explanation is called for.

When Jones says, "Play *X* is better than Play *Y*, and your time would be better spent if you went to Play *Y*," various things may be said in explanation. Jones may point out respects in which viewing *Y* is especially useful. For example, appreciating *Y* may provide a useful introduction to a wide range of great literature. Or viewing *Y* may be very helpful to Smith's preparation of his dissertation on minor dramatists of the 18th century. Or, it may provide historical insights, thus helping Smith to prepare his dissertation on class struggle in the 18th century. In such cases, we might say that the greater value of experiencing *Y* is instrumental rather than aesthetic.

Alternatively, Jones may point to extra-aesthetic elements of great intrinsic value which are likely to accompany Smith's experience of *Y*. *Y* may be likely to remind Smith of some extremely delightful and profound experience which he had. Alternatively again, it may be simply that Smith is not as well equipped to appreciate *X* as he is to appreciate *Y*, so that even though *X* is better it would be "wasted" on him. Yet another alternative is that Smith has been seeing a number of plays of the same general character as *X*, so that *Y*, even though somewhat inferior, would provide some needed variety.

An explanation of any one of these kinds may convince us that indeed Smith would be better off if he saw the inferior play. The explanation in effect points to some factor that outweighs the immediate, purely aesthetic factor in determining what Smith would be advised to do. If no such overriding factor is specified, then it is paradoxical to recommend the inferior work. If Jones says, "Play *X* is better than Play *Y*, and your time would be better spent if you went to Play *Y*," and then adds that the two plays are more or less on a par with regard to factors other than their aesthetic merit, he is talking nonsense.

"Play *X* is better than Play *Y*, and your time would be better spent if you went to Play *X*" seems close to being the very opposite of nonsense. If Smith has been asking, "Why should I go to *X* rather than *Y*?", the answer "Because *X* is a better play" in most circumstances provides all of the reason that one might wish. Again, there are some circumstances in which more discussion would be required. If it is clear that *Y* would give Smith unusually great help in preparing a dissertation, or would remind him of some extremely delightful and profound experience, it may be that "*X* is a better play" will not be a conclusive reason for seeing *X* rather than *Y*. But if there is no factor which makes seeing *Y* especially more advantageous than seeing *X*, "*X* is a better play" would normally be treated as a conclusive reason for seeing *X*.

To point this out is to do what philosophers so often do: remind ourselves of the obvious. A great deal of study has been devoted recently to the arcane and sophisticated functions of critics.[6] The greatest critics surely do a great deal more than tell us which works are good and which are bad. But it is true also that most critics, especially those who write for newspapers, are read for their evaluations of the works that they review. The immediate practical point, for many readers, is that we do not have the time to read every new novel, see every new play, etc. Even if we were condemned to have the time, we would not want such catholic experience. The critic tells us which plays are worth seeing, and which novels are worth reading. That is, he makes an ethical judgment as to how our time probably could be made most valuable. And he does this, ultimately, by saying which works are good and which are bad.

---

[6] The most stimulating attempts are the essays by Isenberg and Margaret MacDonald, both reprinted in ***Aesthetics and Language*** (Oxford, 1959).

Bernard Harrison makes a related point when he says that statements of aesthetic approval refer "to the probabilities of works providing critical development, or repaying sympathetic attempts to understand them."[7] This suggests an important distinction. The newspaper theater critic normally is concerned with how valuable the experience of a play is likely to be on a first viewing. If the time required for a first viewing will be well spent, he recommends the play. A critic writing, as it were, for the ages, may be concerned with how valuable the experience of a play will be at the fifth, sixth, or subsequent viewing. Inexhaustibility is crucial to our conception of what great art is. Harrison points out that we do not commonly reach a point in our appreciation of great works "at which it can be said, 'now the true work has been revealed'. . . . It is still, for example, thought worthwhile to devote books to the study of 'Hamlet' and some of them still succeed in shedding new light on the interpretation of the play."[8]

(2) There is also a heuristic relation between ethical and aesthetic value. This is often disguised by the articulateness of critics, who give ample reasons to support their evaluations of works of art. This creates the impression that they arrived at the evaluations by assembling observations about the works, and then reasoning from these observations to an evaluative conclusion. It also creates the impression that this is an adequate way to arrive at aesthetic evaluations.

I have argued elsewhere that there is a logic of aesthetic argument, but that it is far more complicated than it appears.[9] It is rare that a critic's reasons imply his evaluative conclusion. It then is true, generally speaking, that reasoning to an evaluative conclusion is not by itself an adequate way of evaluating a work of art.

Further it is doubtful that many critics actually work in this way. F. R. Leavis has remarked, "The critic doesn't ask, 'How does this accord with these specifications of goodness in poetry?'; he aims to make fully conscious and articulate the immediate sense of value that places the poem."[10] This suggests that in very many cases the reasons that critics provide to support their evaluations are fully formulated only after the evaluations have been arrived at. This is not to suggest that the reasons are bogus, or lack logical strength: quite to the contrary. But it does suggest that we have to look elsewhere if we wish to find out how in very many cases the critic does arrive at his evaluations.

I shall assume that Leavis is right, and that there is a "sense of value" that the critic tries to articulate. Of what does this "sense of value" consist? There is an obvious candidate, if we draw again on Moore. Implicit throughout *Principia Ethica* is the proposition that we can make judgments of the intrinsic value of our states of mind; and that we do readily, in a variety of circumstances, make such judgments. If we accept this, then we might say that we are immediately aware that certain experiences (e.g., those of long hours spent in drudgery) are of low intrinsic value, and that certain other experiences (e.g., those of high minded friendship) are of high intrinsic value.

Take now the case of a critic scrutinizing a new poem or a new painting. It would be wrong to suggest that his experience is totally unshaped or value-neutral until the moment at which he arrives at an evaluation of the work. He will notice certain features of the work: some of these may well be characteristic of good works, or of bad works. However, this century has accustomed us to exceptions to critical generalizations; so that the critic, unless he is a bit of a dogmatist, will not treat these features as implying a verdict, although he may well treat them as inclining toward a certain kind of verdict. All the while, of course, the critic is having an experience of the work. We may suppose that the critic like other men makes value judgments. He immediately judges experiences of boredom or drudgery to be of low value, and judges experiences which offer (say) both serenity and variety to be valuable. Normally he will place a value also on his experience of the work of art. This, I suggest, is the "sense of value" of which Leavis speaks. If the value is high, the critic may be expected to call the work good. If the value is low, he may be expected to call the work bad.

---

[7] "Some Uses of 'Good' in Criticism," *Mind*, vol. 69 (1960), p. 222.

[8] *Ibid.*, p. 211.

[9] "Reasons in Support of Evaluations of Works of Art," *The Monist*, vol. 50 (1966), pp. 222–236.

[10] *The Common Pursuit* (London, 1952), pp. 213–214.

This suggests that if there is any immediate awareness of value generated by aesthetic experience, the value of which we are immediately aware is ethical. It is the value of the aesthetic experience. If it is high, the aesthetic experience is desirable, worth having, and to be recommended; if it is low, the aesthetic experience is the opposite of these.

This does not mean that we can reduce aesthetic value to ethical value. The relation of the aesthetic value of a work to the ethical values of the experiences of it that various people have is complex and not precisely formulable. We can say that the aesthetic value depends on ethical values, and that we become aware of the aesthetic value by means of awareness of ethical value.[11]

The University of Connecticut

Received April 2, 1971

---

[11] I am endebted to the University of Connecticut Research Foundation for assistance in completing this paper.

American Philosophical Quarterly
Volume 9, Number 3, July 1972

VII. ON AN ARGUMENT FOR TRUTH-FUNCTIONALITY

ROBERT CUMMINS AND DALE GOTTLIEB

I. The Argument

A well-known argument of Quine's[1] is supposed to show that any context of a sentence admitting (without change of truth value) substitution of logically equivalent sentences and coextensive individual descriptions is a truth functional context of that sentence. The argument is as follows. Let 'p' represent any sentence, and let "$C(p)$" represent any sentence containing the sentence represented by 'p'. It is assumed that the sentence represented by "$C(p)$" retains its truth value if (1) the sentence represented by 'p' is replaced by any sentence logically equivalent to it, or if (2) any individual description in the contained sentence is replaced by a coextensive individual description. We notice that regardless of what sentence 'p' happens to represent, it is logically equivalent to

(i) $\hat{x}(x = \varphi \,\&\, p) = \hat{x}(x = \varphi)$.[2]

Hence, we may substitute (i) for 'p' in "$C(p)$" and obtain

(ii) $C(p) \equiv C(\hat{x}(x = \varphi \,\&\, p) = \hat{x}(x = \varphi))$.

If we assume

(iii) $p \equiv q$,

it follows that

(iv) $\hat{x}(x = \varphi \,\&\, p) = \hat{x}(x = \varphi \,\&\, q)$.

Now (iv) enables us to substitute "$\hat{x}(x = \varphi \,\&\, q)$" for "$\hat{x}(x = \varphi \,\&\, p)$" in (ii) yielding,

(v) $C(p) \equiv C(\hat{x}(x = \varphi \,\&\, q) = \hat{x}(x = \varphi))$.

But since whatever sentence 'q' represents, it is logically equivalent to

(vi) $\hat{x}(x = \varphi \,\&\, q) = \hat{x}(x = \varphi)$,

we may replace the occurrence of (vi) in (v) by 'q' obtaining

(vii) $C(p) \equiv C(q)$.

Since the only assumption made about the sentence represented by 'q' is (iii), it is concluded that the context in question is truth functional.

This argument is fallacious. It can be repaired in a way congenial to Quine's philosophical preferences, but at least one recent application of the argument cannot be salvaged in the repaired version.

II. The Fallacy

How is the set abstraction operator to be understood? There are two possibilities. (a) The set abstraction operator is to be defined contextually, perhaps as follows:

$$x \in \hat{x}Fx =_{df} Fx$$
$$\hat{x}Fx = \hat{x}Gx =_{df} (\forall x)\ (x \in \hat{x}Fx \equiv x \in \hat{x}Gx).$$

Combining these definitions we get,

$$\hat{x}Fx = \hat{x}Gx =_{df} (\forall x)\ (Fx \equiv Gx).$$

Now if we adopt this interpretation, the move from (iv) to (v) becomes illegitimate. For although (iv) is true, it does not assert the coextensiveness of two individual descriptions, but rather of two predicates, and we have no provision for the substitution of coextensive *predicates*. (b) The set abstraction operator makes definite descriptions out of predicates (open sentences). Now in first order logic with identity, "$\ldots c \ldots \supset (\exists x)\ (x = c)$" is provable, where '$c$' is any expression which can occupy the position of an individual variable. Thus, (i) logically implies the existence of $\hat{x}(x = \varphi)$ no matter what sentence 'p' represents. But then that sentence is not logically equivalent to the sentence represented by (i), since it is not the case that every sentence implies the existence of $\hat{x}(x = \varphi)$.

At this point it might be objected that every sentence (of a first order theory)[3] *does* imply the existence of $\hat{x}(x = \varphi)$ on the following grounds. First order theories contain axioms of identity (to express uniqueness, functionality, etc.). Given any individual constant '*c*' of such a theory, we have the following proof:

---

[1] W. V. O. Quine, "Reference and Modality," in *From a Logical Point of View* (p. 159), and "Three Grades of Modal Involvement," in *Ways of Paradox*, (pp. 161–162). The argument ultimately goes back to Frege.

[2] 'φ' is here to be construed as a name for the empty set.

$(\forall x)\ (x = x)$ (Axiom of identity)
Hence, $c = c$
Hence, $(\exists x)\ (x = c)$

In our case, we have "$(\exists x)\ (x = \hat{x}(x = \varphi))$" as a theorem. If one regards identity theory as part of logic,[4] then this theorem is guaranteed by logic alone. So "$p \supset (\exists x)\ (x = \hat{x}(x = \varphi)$" will likewise be a theorem of logic.

Something has obviously gone wrong here: Must we say that "Grass is green," or even "There are no classes," logically imply the existence of $\hat{x}(x = \varphi)$? Surely this is going too far. For the objector's argument can be used to "logically guarantee" the existence of anything, since the following is also a proof.

$(\forall x)\ (x = x)$
Hence, Pegasus = Pegasus
Hence, $(\exists x)(x = \text{Pegasus})$

Obviously, the problem is that the adoption of a singular term as part of our vocabulary is not an ontologically neutral matter. To take an expression, e.g., the letter '*c*', as a singular term in the vocabulary of a theory is to be committed to the existence of something named by '*c*'. The axiom of identity, if counted as part of the logic of a first order theory at all, can only be done so in a first order theory without singular terms.

Still, a determined objector might argue as follows. "After all, the formula "$(\exists x)(x = c)$" is *valid*, assuming the ordinary interpretation of '='. So how can its appearance as a theorem of a first order theory be objectionable?"

The answer is that "$(\exists x)(x = c)$" is logically valid — comes out true in every interpretation — only as long as '*c*' has the status of an uninterpreted individual constant. For then, by the definition of an interpretation, we must have a non-empty universe for the quantifier to range over, and '*c*' must be assigned *some thing* in that universe, thus making "$(\exists x)(x = c)$" true. But if '*c*' is a meaningful singular term, then "$(\exists x)(x = c)$" is no longer logically valid, for it asserts the existence of a thing named by that term.

"Well then, can't the argument be salvaged simply by rewriting it using the uninterpreted individual constant '*c*' for 'φ'?" This question reveals a rather subtle misunderstanding of the role of the symbolic expressions in the argumentation of Sect. I. What Sect. I purports to provide is an argument schema. The *claim* made for this schema is that it is valid, i.e., that its instances are good arguments. An instance of a schema is the result of fixing the interpretation of all the uninterpreted symbols in some way or other which is consistent with the syntactic roles played by those symbols. In our case, since the interpretations of such symbols as "$\hat{x}(\quad)$," '=', and '&' are assumed fixed in the usual way, what is required is that '*p*' and '*q*' be interpreted as some particular sentences, "*C*( )" as some particular containing sentence. Now if we replace 'φ' with '*c*', we must ask what role '*c*' is to play. If it is simply an uninterpreted symbol which is not a place-holder for a meaningful expression, then we no longer have a schema instances of which are arguments, but a mixture of prose on the one hand, and symbolic expressions of undetermined import on the other. On the other hand, if, as its syntactic position suggests, '*c*' is treated as a place-holder for singular terms, we shall have an argument schema, but not a valid one. For it is not true that, regardless what sentence '*p*' happens to represent, and what singular term '*c*' happens to represent, '*p*' and "$\hat{x}(x = c \,\&\, p) = \hat{x}(x = c)$" are logically equivalent. For example, "Grass is green," is not logically equivalent to "$\hat{x}(x = \varphi \,\&\, \text{Grass is green}) = \hat{x}(x = \varphi)$." In fact, the suggestion that individual constants of the first order calculus be interpreted as place-holders for singular terms on a par with sentence letters results in the serious anomaly that instances of theorems of the first order calculus may be false. For example, '$(\exists x)(x = \text{Pegasus})$' is a false instance of the theorem "$(\exists x)(x = c)$."

We may see our way to reducing the gross anomaly of getting a falsehood as an instance of a theorem of logic by noticing that even with predicate letters, we must choose our English predicates carefully. "$(Fa \,\&\, a = b) \supset Fb$" is a provable schema, but if we let "John believes —— is tall" go in for 'F ——' we shall have well-known problems on our hands. The moral is that for the purposes of logic, an English predicate is not just a sentence minus some singular terms, but rather an expression that functions referentially as a predicate — i.e., is true or false of objects.[5] We may say similarly in the case of an individual constant that it can be replaced only by an expression which functions referentially as a singular term — i.e., denotes some object. Thus, "Something is Pegasus" will not be regarded as an instance of "$(\exists x)(x=c)$." But "Something is the sun" will be. And this is the lesser anomaly we are left with: Whereas "Something is the sun" is true, it is hardly logically true.

---

[3] By a first order theory, we mean a formal theory with a given interpretation whose underlying logic is standard first order logic. Cf. Alonzo Church, *Introduction to Mathematical Logic*, § 07 and 55. In particular, it is to be emphasized that the logic in question is not a "free logic," since, in such a logic, the following argument is clearly not valid.

[4] We do not wish to commit ourselves for or against such a position. If identity theory is *not* regarded as part of logic, then the objection under consideration is *clearly* without merit. Hence we argue under the opposite assumption.

It emerges, then, that (i)–(vii) cannot be reformulated using an individual constant in place of 'φ'. For either the individual constant is not to be construed as a schematic letter, in which case Sect. I does not provide us with an argument at all, or the individual constant is to be treated as schematic for singular terms, in which case our original point remains: The sentence represented by 'p' will not be logically equivalent to that represented by (i), only the latter implying the existence of $\hat{x}(x = \varphi)$.

III. REPAIRS

Following the hint in (a) of the last section, let us assume that substitution of coextensive predicates in the contained sentence leaves the truth value of the containing sentence untouched. The desired conclusion will then be forthcoming. The repaired argument runs as follows: Regardless of what sentence 'p' happens to represent, it is logically equivalent to

(i) $(\forall x)(p \mathbin{\&} x = x \equiv x = x)$.

Hence we may substitute (i) for 'p' in "$C(p)$" and obtain

(ii) $C(p) \equiv C(\,(\forall x)(p \mathbin{\&} x = x \equiv x = x)\,)$.

If we assume

(iii) $p \equiv q$,

it follows that

(iv) $(\forall x)(p \mathbin{\&} x = x \equiv q \mathbin{\&} x = x)$.

Now (iv) allows us to replace "$p \mathbin{\&} x = x$" by "$q \mathbin{\&} x = x$" in (ii), yielding

(v) $C(p) \equiv C(\,(\forall x)(q \mathbin{\&} x = x \equiv x = x)\,)$.

But since, whatever sentence 'q' represents, it is logically equivalent to

(vi) $(\forall x)\ (q \mathbin{\&} x = x \equiv x = x)$,

we can replace the occurrence of (vi) in (v) by 'q' and obtain

(vii) $C(p) \equiv C(q)$.

This reformulation of the argument we take to be well within the spirit of Quine's often expressed preference for the elimination of singular terms in favor of predicates.

IV. USES OF THE ARGUMENT

In an attempt to show that the logical form of singular causal sentences such as

(1) The short circuit caused the fire.

cannot be given by sentences such as

(2) *The fact that* there was a short circuit *caused it to be the case that* there was a fire.

where the italicized words constitute some sort of sentential connective, Donald Davidson employs a version of the argument under discussion.[6]

---

[5] The same point is made by Massey, *Understanding Symbolic Logic*, pp. 228ff.

[6] Donald Davidson, "Causal Relations," *Journal of Philosophy*, vol. 64 (1967), pp. 694–695. Davidson regards the argument on pages 197–198 of Quine's *Word and Object* as closely related to that of Sect. I. Briefly, Quine is there arguing against a version of modal logic which limits the universe over which the variables range to objects which are not specifiable in ways that fail of necessary equivalence. This limitation is to be accomplished by adopting a postulate, viz.

$$(\exists x)\ (Fx \mathbin{\&} (\forall y)\ (Fy \supset y=x) \mathbin{\&} Gx \mathbin{\&} (\forall y)\ (Gy \supset y=x)\ ) \supset \mathcal{N}(\forall x)\ (Fx \equiv Gx).$$

From this postulate Quine deduces that, for any sentence p, if p then $\mathcal{N}p$. The argument, slightly reformulated, is as follows: Assume p, and let 'a' and 'b' be any singular terms such that $a=b$. We then have

$a=b \mathbin{\&} (\forall y)\ (y=b \supset y=a)$, and

$(a=b \mathbin{\&} p) \mathbin{\&} (\forall y)\ (\,(y=b \mathbin{\&} p) \supset y=a)$.

From the postulate, with "Fx" as "$x=b$" and "Gx" as "$x=b \mathbin{\&} p$," we then have

$\mathcal{N}(\forall x)\ (x=b \equiv (x=b \mathbin{\&} p)\,)$.

Instantiating with respect to 'b' we get

$\mathcal{N}(b=b \equiv (b=b \mathbin{\&} p)\,)$.

By standard modal logic we finish with

$\mathcal{N}(b=b)$, hence

$\mathcal{N}(b=b \mathbin{\&} p)$, hence

$\mathcal{N}p$.

This argument is different from that of Sect. I. In particular, this argument, unlike that of Sect. I, is valid.

> It is obvious that the connective in (2) is not truth-functional, since (2) may change from true to false if the contained sentences are switched. Nevertheless, substitution of singular terms for others with the same extension in sentences like (1) and (2) does not touch their truth value.[7]

Davidson supports this claim with two examples, and then continues,

> Surely also we cannot change the truth value of the likes of (2) by substituting logically equivalent sentences for sentences in it. Thus, (2) retains its truth if for "there was a fire" we substitute the logically equivalent '$\hat{x}(x=x \;\&\; \text{there was a fire}) = \hat{x}(x=x)$'; retains it still if for the left side of this identity we write the coextensive singular term "$\hat{x}(x=x \;\&\; \text{Nero fiddled})$"; and still retains it if we replace "$\hat{x}(x=x \;\&\; \text{Nero fiddled}) = \hat{x}(x=x)$" by the logically equivalent "Nero fiddled." Since the only aspect of "there was a fire" and "Nero fiddled" that matters to this chain of reasoning is the fact of their material equivalence, it appears that our assumed principles have led to the conclusion that the main connective of (2) is, contrary to what we supposed, truth-functional.[8]

The problem, of course, is that "there was a fire" is not logically equivalent to "$\hat{x}(x = x \;\&\; \text{there was a fire}) = \hat{x}(x = x)$," since the latter, but not the former, implies the existence of the universal set. This commitment to the universal set can be avoided if "$\hat{x}(x = x \;\&\; \text{there was a fire}) = \hat{x}(x = x)$" is construed as virtual notation asserting the coextensiveness of two predicates. But this move is of no help to Davidson: Although "substitution of singular terms for others with the same extension in sentences like (1) and (2) does not touch their truth value," substitution of general terms for others with the same extension in sentences like (1) and (2) *does* touch their truth value. To see this we need only reflect that whatever caused it to be the case that John F. Kennedy was born second son to Joseph and Rose Kennedy quite certainly did not cause it to be the case that John F. Kennedy was inaugurated President of the United States in 1960. Yet "was born second son to Joseph and Rose Kennedy" and "was inaugurated President of the United States in 1960" are coextensive predicates.

Davidson remarks that, "Having already seen that the connective of (2) cannot be truth-functional, it is tempting to try to escape the dilemma by tampering with the principle of substitution which led to it."[9] Davidson clearly regards such a strategy as unpromising and undesirable. We agree that tampering with substitutivity of identicals and of logical equivalents in the likes of (2) is undesirable. But these principles lead to no dilemma. Substitutivity of coextensive *predicates* and of logical equivalents in the likes of (2) does lead to a dilemma, but, in this case, tampering is not undesirable, it is mandatory. Substitution of coextensive predicates does not leave the truth-value of the likes of (2) untouched.

Quine makes use of an argument similar to the argument of Sect. I to show that the assumption of transparency for the context "Tom believes that . . ." is tantamount to the assumption of truth-functionality for that context.

> Where 'p' represents a sentence let us write 'δp' (following Kronecker) as short for the description:
>
> the number x such that ($(x=1)$ and p) or ($(x=0)$ and not p).
>
> We may suppose that Tom . . . is enough of a logician to believe a sentence of the form "$\delta p=1$" when and only when he believes the sentence represented by 'p'.[10]

Now suppose Tom believes that Cicero denounced Cataline. Then Tom believes that δ(Cicero denounced Cataline) $= 1$. But, assuming 'p' represents a true sentence, $\delta p = \delta$(Cicero denounced Cataline). Substituting "δp" for "δ(Cicero denounced Cataline)" in "Tom believes that δ (Cicero denounced Cataline)," which is justified by our assumption that the context "Tom believes that . . ." is transparent, we have that Tom believes that $\delta p = 1$. From this, and the assumption that Tom believes sentences of the form "$\delta p = 1$" when and only when he believes the sentence represented by 'p' we have that Tom believes that p.

This argument avoids the difficulties present in the version of Sect. I by requiring, not that 'p' and "$\delta p = 1$" are logically equivalent, but by assuming that Tom believes sentences of the form "$\delta p = 1$" when and only when he believes the sentence represented by 'p'. Now Quine *implies* that Tom does only what any good logician would do and with this we must disagree. For while sentences of the form "$\delta p = 1$" commit one to a number, the sentence p need not, and in general, will not. But, although the rest of us, on reflection, may not believe sentences of the form "$\delta p = 1$" when and only when we believe the sentence p, it is quite possible that *Tom* does. Thus, though somewhat misleading, Quine's argument does establish the desired conclusion, viz., that the context, "x believes that . . ." cannot be generally treated as transparent.

---

[7] *Ibid*, p. 694.

[8] *Ibid*, p. 695.

[9] *Ibid*, p. 695.

[10] W. V. O. Quine, *Word and Object*, *op. cit.*, pp. 148–149.

V. Postscript

Having established the invalidity of Quine's argument for the thesis that the truth-functionality of a context is guaranteed by substutivity of logical equivalents and co-designative singular terms, there remains the question of its truth. Interestingly, what appears to be a counter-example to the thesis is suggested by Davidson's discussion of singular causal statements. Consider the following sentence.

> The fact that Smith fainted caused it to be the case that Jones ran for help.

The contained sentences ("Smith fainted" and "Jones ran for help") may always be replaced by sentences logically equivalent to them without changing the truth value of the entire sentence. Also, the singular terms in the contained sentences may always be replaced by other singular terms having the same referents without changing the truth value of the containing sentence.[11] Yet it is well-known that the context

The fact that . . . caused it to be the case that—.

is not truth functional.

The Johns Hopkins University

Received September 24, 1971

---

[11] Davidson gives some confirming examples in his discussion on pp. 694–695 of "Causal Relations," *Journal of Philosophy*, vol. 64, (1967).

American Philosophical Quarterly
Volume 9, Number 3, July 1972

VIII. WITTGENSTEIN AND THE TRUTH-FUNCTIONALITY THESIS

HUGUES LEBLANC

Wittgenstein claimed in the *Tractatus* that any statement is a truth-function of its components, indeed a truth-function of its elementary components. Few of us today would share either view; too many statements resist the kind of analysis that Wittgenstein envisioned. But the kindred claim that Russell's ramified theory of types (a "model language" to the Wittgenstein of the *Tractatus*) has a *strictly truth-functional interpretation* is sound, and I shall defend it here. I shall go one step farther, and show (in outline form) that —upon adoption of Russell's *Axiom of Reducibility* —the principal interpretation of the theory, though still *truth-functional* in a sense is no longer strictly truth-functional, a result which perhaps justifies Wittgenstein's own uneasiness over the axiom.

To begin with a straightforward example, consider SC (= the sentential calculus); employ, say, '$\sim$' and '$\supset$' as primitive connectives; suppose SC has the same theorems as, say, Church's P_2 in [1]; and, generalizing the notion of a truth-value assignment, understand by a *truth-value function for SC* any function from the wffs (non-atomic as well as atomic) of SC to $\{T, F\}$.

I shall say that SC has a *truth-functional interpretation* if there is a family Σ of truth-value functions for SC such that:

(1) the theorems of SC and the wffs of SC evaluating to T under all members of Σ coincide, and

(2) the truth-values that negations and conditionals assume under members of Σ depend on the truth-values of their components, and on those truth-values only.

And I shall say that SC has a *strictly truth-functional interpretation* if there is a family Σ of truth-value functions for SC meeting conditions (1)–(2) *plus* the following condition:

(3) members of Σ that agree on the atomic wffs of SC agree on the non-atomic ones as well.[1]

Now let Σ' consist of every truth-value function α for SC such that: (a) $\alpha(\sim A) = T$ if and only if $\alpha(A) = F$, and (b) $\alpha(A \supset B) = T$ if and only if $\alpha(A) = F$ or $\alpha(B) = T$. It is easily verified that Σ' meets all three of conditions (1)–(3), and hence that SC has a strictly truth-functional interpretation, a result which owes much to Wittgenstein. (The result can be considerably strengthened. Require in my account of truth-functionality that all truth-value assignments to the atomic wffs of SC figure among the members of Σ (condition (0)). It can be verified that only one family of truth-value functions for SC meets conditions (0)–(2): the family Σ', and hence that all truth-functional interpretations of SC are strictly truth-functional.)

Things get slightly more complicated when you come to QC^1, the first-order quantificational calculus. The truth-value of a negation $\sim A$ will again hinge on that of its component A, and the truth-value of a conditional $A \supset B$ hinge on those of its components A and B. But what will the truth-value of a quantification $(\forall X)A$ hinge on?[2] Wittgenstein supplied the answer in the *Tractatus:* the truth-values of its *substitution instances*.

To handle matters of substitution in the simplest possible way, suppose that in keeping with recent practice we use *two* runs of individual signs: one run (say, the letters '*x*', '*y*', '*z*', etc.) serving in effect as *bound* individual variables, and called *individual variables*; the other (say, the letters '*a*', '*b*', '*c*', etc.) serving as *free* individual variables, and called *individual parameters*. And suppose that formulas of QC^1 in which identical quantifiers overlap are not counted well-formed. The substitution instances of a quantification $(\forall X)A$ of QC^1

---

[1] So that the truth-values of non-atomic wffs of SC depend entirely on those of atomic ones. See Appendix II for a more formal account of things.

[2] I assume throughout that just '$\forall$' serves as primitive quantifier letter, and that '$\exists$' is defined in the customary manner.

will then be all results of putting an individual parameter of QC^1 for each occurrence of X in A; and, in particular, the substitution instances of "$(\forall x)f(x)$" will be—as they were for Wittgenstein —"$f(a)$", "$f(b)$", "$f(c)$", etc.

These preliminaries once disposed of, read "QC^1" for "SC" in my accounts of truth-functionality and strict truth-functionality; in condition (2) require the truth-values that quantifications assume under members of Σ to depend on the truth-values of their substitution instances, and on those truth-values only; and let Σ' consist of every truth-value function α for QC^1 such that:[3] (a) $\alpha(\sim A) = T$ if and only if $\alpha(A) = F$, (b) $\alpha(A \supset B) = T$ if and only if $\alpha(A) = F$ or $\alpha(B) = T$, and (c) $\alpha((\forall X)A) = T$ if and only if $\alpha(A^*) = T$ for every substitution instance A^* of $(\forall X)A$. It can be verified that Σ' meets all three of conditions (1)–(3), and hence that QC^1 has a strictly truth-functional interpretation, a result to which Wittgenstein contributed some.[4] (Again the result can be considerably strengthened. Require that all truth-value assignments to the atomic wffs of QC^1 figure among the members of Σ (condition (0)). It can be shown that only one family of truth-value functions for QC^1 meets conditions (0)–(2): the family Σ', and hence that all truth-functional interpretations of QC^1 are strictly truth-functional.)

Now let us turn to QC^2 (= the second-order quantificational calculus), and in particular to QC$^{2/\infty}$ (= the ramified second-order quantificational calculus); and let us employ both predicate variables (= *bound* predicate variables) and predicate parameters (= *free* predicate variables).[5] As the reader knows, formulas of the sort $(\forall F^d)A$, where F^d is a predicate variable of degree d ($d \geq 0$) are acknowledged as well-formed both in QC^2 and in QC$^{2/\infty}$. However, whereas QC^2 has a single run of predicate variables, QC$^{2/\infty}$ has one for each so-called *level* ℓ ($\ell = 1, 2, 3, \ldots$); and, whereas QC^2 has a single run of predicate parameters, QC$^{2/\infty}$ has one for each level ℓ. To appreciate Russell's distinction between, say, predicates of level 1 and predicates of level 2, compare "Napoleon had sangfroid on the battlefield" and "Napoleon had all the qualities of a great general"[6]. The predicate "had all the qualities of a great general" could be regarded as summary for such predicates of level 1 as "had sangfroid on the battlefield," "knew how to pen a victory bulletin," etc., and hence could be adjudged of level 2.

Needed at this point are a substitution convention for predicate variables and a means of assessing the level of a wff of QC$^{2/\infty}$. I shall relegate the first to Appendix I, and ask the reader to think of

$$A(B/F^{d/\ell}(X_1, X_2, \ldots, X_d)),$$

where A is a wff or a qwff,[7] B is a wff, $F^{d/\ell}$ is a predicate variable of degree d and level ℓ, and $X_1, X_2, \ldots,$ and X_d are distinct individual parameters of QC$^{2/\infty}$, as behaving in effect like Church's

$$S_B^{F^{d/\ell}(X_1, X_2, \ldots, X_d)}A|$$

in [1]. As to the second desideratum, B being a wff of QC$^{2/\infty}$, let ℓ_P ('P' for 'parameter') be 0 if no predicate parameter of QC$^{2/\infty}$ occurs in B; let ℓ_P be 1 if at least one predicate parameter of level 1 occurs in B but none of level higher than 1 does; let ℓ_P be 2 if at least one predicate parameter of level 2 occurs in B but none of level higher than 2 does; etc. And define ℓ_V ('V' for 'variable') in the same manner, but with the word 'variable' doing duty for 'parameter'. The *level of B* will then be $\max(\ell_P, \ell_V + 1)$. This done, you may take the substitution instances of a quantification $(\forall F^{d/\ell})A$ of QC$^{2/\infty}$ to be all wffs of QC$^{2/\infty}$ of the sort $A(B/F^{d/\ell}(X_1, X_2, \ldots, X_d))$, and take its *predicative* substitution instances to be all wffs of QC$^{2/\infty}$ of that sort in which B is of level ℓ or less.

The distinction just drawn between the substitution instances of $(\forall F^{d/\ell})A$ in general, and its predicative ones in particular, is crucial. Indeed, call a substitution instance A^* of a quantification $(\forall V)A$ *permissible* if

$$(\forall V)A \supset A^*$$

is provable in QC$^{2/\infty}$. Without the *Axiom of Reducibility*, $(\forall F^{d/\ell})A \supset A^*$—though provable in QC$^{2/\infty}$ for every *predicative* substitution instance

---

[3] I.e., of every function α from the wffs (non-atomic as well as atomic of) QC^1 to $\{T, F\}$ such that . . . . I assume in what follows that QC^1 has in effect the same theorems as Church's F^1 in (1).

[4] Ironically, Wittgenstein came to abjure the *substitution interpretation* of the quantifiers. Moore reports in [4], p. 297: "He said that there was a temptation, to which he had yielded in the *Tractatus*, to say that $(x).fx$ is identical with the logical sum '$fa.fb.fc\ldots$', and $(\exists x).fx$ identical with the logical sum '$fa \vee fb \vee fc \ldots$'; but that this was in both cases a mistake."

[5] Sentence variables (parameters) can be thought of as predicate variables (parameters) of degree 0. So nothing is said about them in the text.

[6] The example stems from Russell.

[7] Qwffs (=quasi-well-formed formulas) are like wffs except for exhibiting variables at places where wffs would exhibit parameters.

A^* of $(\forall F^{d/l})A$—fails for many an impredicative one. On the other hand, Russell's axiom (which guarantees that for any property or relation of level higher than 1 there exists a co-extensive one of level 1) permits proof of $(\forall F^{d/l})A \supset A^*$ for *every* substitution instance A^* of $(\forall F^{d/l})A$.[8] So, in one case only some of the substitution instances of $(\forall F^{d/l})A$ turn out permissible, whereas in the other they all do.

Now read "$QC^{2/\infty}$" for "SC" in my accounts of truth-functionality and strict truth-functionality; in (2) require the truth-values that quantifications assume under members of Σ to depend on the truth-values of their permissible substitution instances, and on those truth-values only; let Σ'_1 consist of every truth-value function α for $QC^{2/\infty}$ such that (a) $\alpha(\sim A) = T$ if and only if $\alpha(A) = F$, (b) $\alpha(A \supset B) = T$ if and only if $\alpha(A) = F$ or $\alpha(B) = T$, (c) $\alpha((\forall X)A) = T$ if and only if $\alpha(A^*) = T$ for every instance A^* of $(\forall X)A$, and (d) $\alpha((\forall F^{d/l})A) = T$ if and only if $\alpha(A^*) = T$ for every predicative substitution instance A^* of $(\forall F^{d/l})A$; and let Σ'_2 be similarly defined, but with (d) now reading: $\alpha((\forall F^{d/l})A) = T$ if and only if $\alpha(A^*) = T$ for every substitution instance A^* of $(\forall F^{d/l})A$.

It is easily verified that *in the absence of the Axiom of Reducibility* Σ'_1 meets conditions (1)–(2), and *upon the adoption of the axiom* so does Σ'_2. In either case, therefore, $QC^{2/\infty}$ has a strictly truth-functional interpretation. And, aided by George Weaver, I have verified that Σ'_1 meets condition (3) as well.(9) However, Robert K. Meyer and myself have isolated two members of Σ'_2 which agree on all the atomic wffs of $QC^{2/\infty}$ and yet disagree on the wff

$$(\exists f)\,(\exists x)\,(\exists y)\,(f(x)\ \&\ \sim f(y)).^{10}$$

So *in the absence of the Axiom of Reducibility* the main interpretation of $QC^{2/\infty}$, to wit: Σ'_1, *is* strictly truth-functional;[11] but *upon adoption of the axiom* its main interpretation, now Σ'_2, is *not* strictly truth-functional.[12]

Like results hold for $QC^{\infty/\infty}$, Russell's ramified theory of types, and the full *Axiom of Reducibility*: in the absence of that axiom $QC^{\infty/\infty}$ *has* a strictly truth-functional interpretation, but upon adoption of the axiom its main interpretation is *not* strictly truth-functional.

Wittgenstein had misgivings of his own over Russell's axiom, which I find hard to evaluate. The results in the last paragraph may be more to the point, given Wittgenstein's advocacy of strict truth-functionality.

Temple University

Received March 15, 1971

References

[1] Church, A., *Introduction to Mathematical Logic, Volume I*, Princeton, 1956.

[2] Leblanc, H., and Meyer, R. K., "Truth-value Semantics for the Theory of Types," in *Philosophical Problems in Logic Some Recent Developments*, Dordrecht-Holland, 1970.

[3] Leblanc, H., and Weaver, G., "Truth-functionality and the Ramified Theory of Types," in *Proceedings of the Temple University Conference on Formal Semantics*, Amsterdam, forthcoming.

[4] Moore, G. E., *Philosophical Papers*, (London, 1959).

---

[8] The *Axiom of Reducibility*, as tailored to suit $QC^{2/\infty}$, reads:

$(\exists F^{d/1})\,(\forall X_1)\,(\forall X_2)\ldots(\forall X_d)\,(F^{d/1}(X_1,X_2,\ldots,X_d) \equiv \mathbf{F}^{d/l}.(X_1,X_2,\ldots,X_d))$, where $\mathbf{F}^{d/l}$ is a predicate parameter of $QC^{2/\infty}$ of degree d and level l ($d \geq$ and $l > 1$), and $X_1, X_2, \ldots$, and X_d are distinct individual variables of $QC^{2/\infty}$. In the absence of the axiom $QC^{2/\infty}$ is presumed here to have in effect the same theorems as $F^{2/\infty}_2$ in [1].

[9] For full details see [3].

[10] For full details see [2].

[11] There is some slippage, though: the truth value of a wff of SC depends exclusively upon the truth values of its atomic *components*, that of a wff of QC^1 depends upon the truth values of its atomic *subformulas* (not all of which need be components of the wff), and that of a wff of $QC^{2/\infty}$ generally depends upon the truth values of *all* the atomic wffs of $QC^{2/\infty}$.

[12] Whether in the absence of the *Axiom of Reducibility* $QC^{2/\infty}$ has any truth functional interpretation which is *not* strictly truth functional, and whether upon adoption of the axiom $QC^{2/\infty}$ has any truth functional interpretation which—unlike its main one—*is* strictly truth functional, are still open questions. I would conjecture that the answer is *No* in both cases.

APPENDIX I[13]

SUBSTITUTION CONVENTION FOR PREDICATE VARIABLES

(1) Let A be a wff or qwff of $QC^{2/\infty}$; let X_1, X_2, . . ., and X_d ($d \geq 0$) be distinct individual parameters of $QC^{2/\infty}$; and let I_1, I_2, . . ., and I_d be (not necessarily distinct) individual signs of $QC^{2/\infty}$. Then $A(I_1, I_2, \ldots, I_d/X_1, X_2, \ldots, X_d)$ is to be the result of simultaneously putting I_1 everywhere in A for X_1, I_2 for X_2, . . ., and I_d for X_d.

(2) Let A be a wff or qwff of $QC^{2/\infty}$; let $F^{d/\ell}$ be a predicate variable of $QC^{2/\infty}$ of degree d and level ℓ; in case $d > 0$, let X_1, X_2, . . ., and X_d be distinct individual parameters of $QC^{2/\infty}$; let B be a wff of $QC^{2/\infty}$; and let $F^{d/\ell}(I_{1_1}, I_{2_1}, \ldots, I_{d_1})$, $F^{d/\ell}(I_{1_2}, I_{2_2}, \ldots, I_{d_2})$, . . ., and $F^{d/\ell}(I_{1_k}, I_{2_k}, \ldots, I_{d_k})$ ($k \geq 0$) be all the (distinct) components of A of the sort $F^{d/\ell}(I_1, I_2, \ldots, I_d)$.

Case 1: $k = 0$. Then $A(B/F^{d/\ell}(X_1, X_2, \ldots, X_d))$ is to be A.

Case 2: $k \geq 0$. (a) If $F^{d/\ell}$ occurs in a component of A of the sort $(\forall V)A'$ and B has a component of the sort $(\forall V)B'$ (V the same variable in both cases), $A(B/F^{d/\ell}(X_1, X_2, \ldots, X_d))$ is to be the result of putting the alphabetically earliest predicate parameter of $QC^{2/\infty}$ of the same degree and level as $F^{d/\ell}$ everywhere in A for $F^{d/\ell}$. (b) Otherwise, $A(B/F^{d/\ell}(X_1, X_2, \ldots, X_d))$ is to be the result of putting $B(I_{1_i}, I_{2_i}, \ldots, I_{d_i}/X_1, X_2, \ldots, X_d)$ everywhere in A for $F^{d/\ell}(I_{1_i}, I_{2_i}, \ldots, I_{d_i})$, this for each i from 1 through k.

APPENDIX II

TRUTH-FUNCTIONALITY AND STRICT TRUTH-FUNCTIONALITY

Note: In the definitions that follow, C may be SC (in which case (2.5) drops out), QC^1, or $QC^{2/\infty}$ (with or without the Axiom of Reducibility); and a *permissible instantial function* for (the quantifications of) C is to be any function that pairs with each quantification $(\forall V)A$ of C a set of permissible substitution instances of $(\forall V)A$.

A calculus C is said *to have a truth-functional interpretation* if there is a family Σ of truth-value functions for C such that:

(1) for any wff A of C, A is provable in C if and only if $\alpha(A) = T$ for every member α of Σ,

(2.1) for any two members α and α' of Σ and any negation $\sim A$ of C, if $\alpha(A) = \alpha'(A)$, then $\alpha(\sim A) = \alpha'(\sim A)$,

(2.2) for any member α of Σ and any two negations $\sim A$ and $\sim A'$ of C, if $\alpha(A) = \alpha(A')$, then $\alpha(\sim A) = \alpha(\sim A')$,

(2.3) for any two members α and α' of Σ and any conditional $A \supset B$ of C, if $\alpha(A) = \alpha'(A)$ and $\alpha(B) = \alpha'(B)$, then $\alpha(A \supset B) = \alpha'(A \supset B)$,

(2.4) for any member α of Σ and any two conditionals $A \supset B$ and $A' \supset B'$ of C, if $\alpha(A) = \alpha(A')$ and $\alpha(B') = \alpha(B')$, then $\alpha(A \supset B) = \alpha(A' \supset B')$, and

(2.5) there is a permissible instantial function I for C such that:

(2.5.1) for any two members α and α' of Σ and any quantification $(\forall V)A$ of C, if $\alpha(A^*) = \alpha'(A^*)$ for every member A^* of $I((\forall V)A)$, then $\alpha((\forall V)A) = \alpha'((\forall V)A)$,

(2.5.2) for any member α of Σ, any two quantifications $(\forall V)A$ and $(\forall V')A'$ (V and V' both individual variables or both predicate variables) of C such that $I((\forall V)A)$ is at least as large as $I((\forall V')A')$, and any function f from $I((\forall V)A)$ onto $I((\forall V')A')$, if $\alpha(A^*) = \alpha(f(A^*))$ for every member A^* of $I((\forall V)A)$, then $\alpha((\forall V)A) = \alpha((\forall V')A')$, and

(2.5.3) for any member α of Σ, any two quantifications $(\forall V)A$ and $(\forall V')A'$ (V and V' as before) of C such that $I((\forall V)A)$ is smaller than $I((\forall V')A')$, and any function f from $I((\forall V')A'$ onto $I((\forall V)A)$, if $\alpha(A'^*) = \alpha(f(A'^*))$ for every member A'^* of $I((\forall V')A')$, then $\alpha((\forall V)A) = \alpha((\forall V')A'$.

And C is said *to have a strictly truth-functional interpretation* if there is a family Σ of truth-value functions for C that meets (1)–(2.5) *plus* the following condition:

(3) for any two members α and α' of Σ if $\alpha(A) = \alpha'(A)$ for every atomic wff A of C, then $\alpha(A) = \alpha'(A)$ for every non-atomic wff A of C.[14]

---

[13] In what follows 'I' ranges over individual variables and parameters. All other syntactical variables have already turned up in the main text.

[14] The paper (which presupposes familiarity with Wittgenstein's *Tractacus*) was read at the 1971 Meeting of the American Philosophical Association, Pacific Division. Thanks are due to Paul Benacerraf for his invaluable comments, and to George Weaver who helped to devise the criterion of truth functionality in Appendix II and prove that $QC^{2/\infty}$ has a strictly truth-functional interpretation.

BOOKS RECEIVED

Louis Barjon, *Le Combat De Pierre Teilhard De Chardin* (Les Presses De l'Universite Laval: Quebec, 1971), 312 pp. $7.50

Norman Bowie, *Distributive Justice* (The University of Massachussetts Press: Lawrence, Mass., 1971), 148 pp. $9.00

Malcolm Brown (ed.), *Plato's Meno* (The Bobbs-Merrill Company, Inc.: New York, 1971), 314 pp. $3.95

Steven M. Cahn, *The Philosophical Foundations of Education* (Harper & Row: New York, 1970), 433 pp. $6.95

Steven M. Cahn, *A New Introduction to Philosophy* (Harper & Row Publishers: New York, 1971), 540 pp.

Milic Capek, *Bergson and Modern Physics* (D. Reidel Publishing Company: Dordrecht, 1971), 414 pp., in *Boston Studies in the Philosophy of Science*, Vol. VII.

Fernand Couturier, *Monde et etre chez Heidegger* (Les Presses de l'universite de Montreal, 1971), 584 pp.

Forest K. Davis, *Return From Enlightenment* (Adamant Press: Vermont, 1971), 160 pp.

David H. DeGrood, *et. al.* (eds.), *Radical Currents in Contemporary Philosophy* (Warren H. Green, Inc.: St. Louis, 1971), 277 pp.

Jonn M. Dorsey, *Psychology of Emotion* (Center for Health Education: Michigan, 1971), 174 pp.

John M. Dorsey, *Psychology of Language* (Center for Health Education: Detroit, 971), 145 pp. $6.95.

Rem Blanchard Edwards, *Freedom, Responsibility and Obligation* (Martinus Nijhoff: The Hague, 1971), 125 pp.

Risieri Frondizi, *What is Value?* (The Open Court Publishing Co.: Illinois, 1971), 169 pp.

Samuel Gorovitz (ed.), *Mill: Utilitarianism* (The Bobbs-Merrill Company, Inc.: New York, 1971), 410 pp.

Arne E. Hermann, *Principles of Culture and the Crisis of Modern Civilization* (Vantage Press: New York, 1971), 300 pp. $4.95.

Edward Kent (ed.), *Revolution and the Rule of Law* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1971), 181 pp. $6.00

Carl R. Kordig, *The Justification of Scientific Change* (D. Reidel Publishing Co: Dordrecht, 1971), 119 pp.

Hans Lenk, *Philosophie im technologischen Zeitalter* (Verlag W. Kohlhammer: Stuttgart, 1971), 174 pp.

A. Lichtigfeld, *Aspects of Jaspers' Philosophy* (University of South Africa: South Africa, 1971), 119 pp.

Kuno Lorenz, *Elemente der Sprachkritik* (Suhrkamp Verlag: Frankfurt, 1970), 268 pp.

J. Mansfeld, *The Pseudo-Hippocratic Tract Peri Hebdomadoon Ch. I–II and Green Philosophy* (Royal Vangorcum Ltd.: The Netherlands, 1971), 271 pp.

D. C. Mathur, *Naturalistic Philosophies of Experience* (Warren H. Green, Inc.: Missouri, 1971), 158 pp.

R. A. McConnell, *Curriculum Guide* (Simon & Schuster, Inc.: New York, 1971), 128 pp. $1.95

Robert Middlekauff, *The Mathers* (Oxford University Press: New York, 1971), 440 pp.

K. Bruce Miller, *Ideology and Moral Philosophy* (Humanities Press: New York, 1971), 200 pp. $7.50

Jürgen Mittelstrass, *Neuzeit und Aufklärung* (Walter De Gruyter: New York, 1971), 651 pp.

Henri Prat, *l'Espace multidimensionnel* (Les Presses de l'universite de Montreal: Montreal, 1971), 156 pp.

David M. Rosenthal (ed.), *Materialism and The Mind-Body Problem* (Prentice-Hall Inc.: Englewood Cliffs, N.J., 1971), 242 pp.

George Santayana, *Lotze's System of Philosophy* (Indiana University Press: Bloomington, 1971), 274 pp.

Richard Schacht, *Alienation* (Doubleday & Co.: New York, 1971), 320 pp. $1.95

Charles M. Sherover, *Heidegger, Kant & Time* (Indiana University Press: Bloomington, 1971), 322 pp.

Samuel Enoch Stumpf, *Philosophy: History and Problems* (McGraw-Hill Book Company: New York, 1971), 368 pp.

Robert S. Summers (ed.), *More Essays in Legal Philosophy* (University of California Press: Berkeley, 1971), 161 pp. $6.50

Frank A. Tillman and Steven M. Cahn, *Philosophy of Art and Aesthetics* (Harper & Row: New York, 1969), 791 pp. $12.95

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences

Edited by Arne Naess and Alastair Hannay

Numbers 1–2 *Summer 1972* *Volume 15*

Editorial Statement	
Insanity and Responsibility	Herbert Fingarette
Mental Structure and Self-Consciousness	Brian O'Shaughnessy
Spinoza and Mental Health	Paul Wienpahl
Towards an Ecology of Consciousness	Frank Barron
Schizophrenia and Perception: a Critique of the Liberal Theory of Externality	James M. Glass
The Validity of Psychotherapy	B. A. Farrell
Review Discussions	
Wollheim on Freud (Richard Wollheim: *Freud*)	Frank Cioffi
The Politics of Mental Illness (Thomas Szasz: *The Manufacture of Madness*)	Ronald B. de Sousa
Illness and Practical Reasoning (David Robinson: *The Process of Becoming Ill*)	Maurice Roche
Mental Illness and the *Lebenswelt* (Erwin W. Straus, Maurice Natanson, and Henri Ey: *Psychiatry and Philosophy*)	Alastair Hannay

Published quarterly by Universitetsforlaget P.O. Box 307, Blindern, Oslo 3, Norway
Annual Subscription: $10.00, £3.30 inc., postage

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Volume 23, Number 1 February 1972

Articles
J. J. Clarke Turing Machines and the Mind-Body Problem
Michael R. Gardner Quantum-Theoretical Realism: Popper and Einstein v. Kochen and Specker
Kenneth S. Friedman Empirical Simplicity as Testability
Alan R. White The Propensity Theory of Probability

Discussions
Jerzy Giedymin On Quine's Epistemological Naturalism
Errol E. Harris Epicyclic Popperism

Review Article
David Miller Back to Aristotle? (Review of R. Harré *The Principles of Scientific Thinking*)

Reviews
D. J. O'Connor on Gerd Buchdahl *Metaphysics and the Philosophy of Science*, A. Rupert Hall on Robert E. Butts and John W. Davis (*eds.*) *The Methodological Heritage of Newton*, J. W. N. Watkins on Edward Erwin *The Concept of Meaninglessness*

CAMBRIDGE UNIVERSITY PRESS
Bentley House, 200 Euston Road, London NW1 2DB
American Branch: 32 East 57th Street, New York, N.Y. 10022
Single issues £1.50 net in U.K. ($4.75 in U.S.A.)
Subscription price £4.50 net in U.K. ($14.50 in U.S.A.)

JOURNAL OF PHILOSOPHICAL LOGIC

Editor-in-Chief: Bas C. Van Fraassen, University of Toronto, Toronto, Ont., Canada.

Editorial Board: A. R. Anderson, Pittsburgh, Pa., U.S.A.; L. Aqvist, Uppsala, Sweden; N. D. Belnap, Jr., Pittsburgh, Pa., U.S.A.; M. J. Cresswell, Wellington, New Zealand; J. E. Fenstad, Oslo, Norway; F. Fitch, New Haven, Conn., U.S.A.; D. Føllesdal, Oslo, Norway; P. T. Geach, Leeds, England; C. Hamblin, Kensington, N.S.W., Australia; H. G. Herzberger, Bombay, India; J. Hintikka, Stanford, Calif., U.S.A.; W. Hodges, London, England; S. Kanger, Uppsala, Sweden; D. Kaplan, Los Angeles, Calif., U.S.A.; S. Kripke, New York, N.Y., U.S.A.; K. Lambert, Irvine, Calif., U.S.A.; R. B. Marcus, Chicago, Ill., U.S.A.; G. J. Massey, Pittsburgh, Pa., U.S.A.; R. Montague, Los Angeles, Calif., U.S.A.; D. Prawitz, Stockholm, Sweden; N. Rescher, Pittsburgh, Pa., U.S.A.; A. Robinson, Jerusalem, Israel; W. C. Salmon, Bloomington, Ind., U.S.A.; D. Scott, Princeton, N.J., U.S.A.; R. F. Strawson, Oxford, England; P. Suppes, Stanford, Calif., U.S.A.; R. H. Thomason, New Haven, Conn., U.S.A.; G. H. von Wright, Helsinki, Finland.

Contents of Volume 1, No. 1

Editorial—M. J. Cresswell / Intensional Logics and Logical Truth—Michael Friedman and Clark Glymour / Of Quanta Had Logic—Dov M. Gabbay / A General Filtration Method for Modal Logics—Dov M. Gabbay / Tense Systems with Discrete Moments of Time, Part I—Jaakko Hintikka / Constituents and Finite Identifiability—Richard Routley and Robert K. Meyer / The Semantics of Entailment II—C. L. Hamblin / Quandaries and the Logic of Rules—C. Daniels / Reference and Singular Referring Terms.

Subscription price per volume of 500 pages Dfl. 100,—(US $32.50) including postage.

One volume is published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY
P.O. Box 17 / Dordrecht - Holland

NOÛS

Nihil philosophici a nobis alienum putamus
Volume VI Number 2 May 1972

CONTENTS

John Wallace/*Belief and Satisfaction*
Thomas Schwartz/*Rationality and the Myth of the Maximum*
Richard Sharvy/Euthyphro 9d - 11b: *Analysis and Definition in Plato and Others*

Symposium on Properties and Relations
Milton Fisk/*Relatedness Without Relations*
Reinhardt Grossmann/*Russell's Paradox and Complex Properties*
Nino B. Cocchiarella/*Properties as Individuals in Formal Ontology*
Robert K. Meyer/*Identity in Cocchiarella's T**

Publications Received

PRINTED IN ENGLAND
by C. Tinling & Co. Ltd., Liverpool, London and Prescot

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, $6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, $6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, $6.00.

No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, $6.00.

No. 5. DESCARTES' THEORY OF NATURE, by James Collins. 1971, $6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Robert Butts
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Arthur C. Danto
James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
R. M. Hare
Rom Harré
Carl G. Hempel
Raymond Klibansky
Hugues Leblanc
Ernan McMullin
Gerald J. Massey
Richard Rorty
George A. Schrader
Wilfrid Sellars
Alexander Sesonske
Richard Taylor
Manley H. Thompson, Jr.
John W. Yolton

VOLUME 9/NUMBER 4 OCTOBER, 1972

CONTENTS

I. K. W. RANKIN: *The Non-Causal Self-Fulfillment of Intention* 279
II. RAYMOND MARTIN: *On Weighting Causes* 291
III. JOHN KEKES: *Fallibilism and Rationality* 301
IV. JOHN LACHS: *Fichte's Idealism* . . 311
V. PATRICK MCKEE: *Non-Conscious Seeing* 319
VI. BARUCH BRODY: *Locke on the Identity of Persons* 327
VII. ROBERT F. LADENSON: *Legitimate Authority* 335
VIII. EDWARD QUEST: *Why Is It Difficult to Deserve Praise?* 343
IX. ALASTAIR MCKINNON: "*Existence*" *in* "*The Existence of God*" . . . 349
X. A. AARON SNYDER: *The Paradox of Determinism* 353
Books Received 357
Corrigenda 359

PUBLISHED IN OXFORD BY BASIL BLACKWELL

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost $14.00, increasing by $1.40 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost $11.20, increasing by $1.40 for each additional 4 pages. Covers will be provided for offprints at a cost of $6 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

American Philosophical Quarterly
Volume 9, Number 4, October 197

I. THE NON-CAUSAL SELF-FULFILLMENT OF INTENTION

K. W. RANKIN

I. Why are Intentions, if not Causes, so Like Them?

APART from the fact that what we do often depends on what we intend, intentions are conspicuously quasi-causal in at least one other way.

An explicit intention antedates the action intended. This, of course, is clearest where the intention has been formed early enough to be revoked, or forgotten, before the time for fulfillment comes round. But even where it persists right up to, or occurs just at, the executive moment of action, it precedes in the sense that it is there at the moment when the action takes off. Some actions, no doubt, are intentional without being intended. Though in no way premeditated, their performance is under the control of the agent's intelligence. A casual contribution to a conversation is, for instance, usually of this kind. More often than not we do not think what to say before saying it. Consequently, it is sometimes inferred that intentionality without premeditation—and even what is essential to intentionality as such—consists of nothing that antedates the intentional action. But, though I believe the inference to be hasty, I shall set it temporarily to one side, by confining myself initially to intentional actions which are intentional because intended.

Now, in the present context, a cause is taken as in some sense a crucial component of a set of states or events which, relative to natural laws, are together a sufficient condition of whatever is caused. Like intentions, such causes quite often precede, though sometimes they may coincide in time with, what depends upon them. Should intentions, then, be classified as causes? And if not, why are they so cause-like? Both of these questions are the concern of what follows.

II. Twelve Clues

1. Typically a categorical or unqualified intention is expressed by an anticipatory form of utterance such as "I am going to *X*," or "I shall (will) *X*," or "I am about to *X*."[1] It can also be expressed in a non-anticipatory way by some verb of volition as in "I intend to *X*," or "I have decided to *X*." But the latter are indirect. They leave inexplicit the important difference between the agent's attitude at the time toward his intended action and the attitude of someone else who reports the agent's intention in some such formula as "He intends to *X*."

An epistemic analogy may show more clearly why.

The three kinds of sentence in question

(S1) "I intend to *X*,"
(S2) "He intends to *X*,"
(S3) "I am about to *X*,"

compare respectively in function to

(S1′) "I believe Richard Hughes was born in 1900,"
(S2′) "He believes Richard Hughes was born in 1900,"
(S3′) "Richard Hughes was born in 1900."

On some occasions the verb "believe" as in (S1′) has a so-called parenthetical use we shall set to one side. On other occasions it simply ascribes a specific belief-attitude to the speaker. (S1′) then ascribes what (S2′) would ascribe, if used by someone else to refer on the same occasion to the speaker of (S1′).

But (S2′) conveys accurately what its speaker means only where his assumption is that the subject of the ascription has a conviction which could be accurately expressed by (S3′). Should the subject have any reservation about what (S3′) states, then (S2′) would convey its speaker's meaning inaccurately. A sentence of the form "He is inclined to believe . . ." would have been better chosen. Consequently, since *ex hypothesi* (S1′) and (S2′) ascribe the same belief-attitude to the same person, (S3′) expresses the belief-attitude which (S1′) ascribes.

---

[1] Perhaps we should distinguish present-tense anticipatory forms, like "I am going to *X*," from past-tense anticipatory forms, like "I was going to *X*." A peculiarity of the latter is that it in no way commits its utterer to a belief in the subsequent occurrence of the action specified, whereas the former does. Throughout this paper, with the present exception, I identify anticipatory forms of utterance exclusively with present-tense anticipatory forms.

Now this illustration of the distinction between what a sentence ascribes, or reports, and what it expresses helps us to differentiate the respective roles of (S1), (S2), and (S3). (S1) and (S2) may both ascribe the same intention to the same person. In that case (S3) expresses the intention both ascribe, provided it is uttered by the speaker of (S1). His linguistic behavior would be inconsistent if, upon uttering (S1), he were to add "But due to circumstances outside my control I am not about to *X*." It would be as incongruous as that of a speaker of (S1′) who denies what (S3′) conveys.

As already noted, however, when (S1) and (S2) convey the same thing, (S1) gives an *indirect* expression to its speaker's intention. Its inclusion of "I" indicates that the person to whom it ascribes the intention is the utterer, and hence that the utterer also subscribes to the expression of that intention in (S3). Similarly when (S1′) and (S2′) convey the same thing, (S1′) gives indirect expression to its speaker's belief. Its inclusion of "I" likewise indicates that the person to whom it ascribes the belief is the utterer, and hence that the utterer also subscribes to the expression of the belief in (S3′). But in neither case can what the "I" indicates be identified with any part of what the sentence including it states. Otherwise we would have to abandon our hypothesis that what each sentence and its third person partner states or conveys is identical.[2] Consequently, the expression of the intention, in the one case, and of the belief, in the other, is only indirect.

2. When *X*-ing is possible, and someone's intention to *X* is formed in all cognizance of the circumstances which make it possible for him, his intention is fully appropriate. But all the same it would seem odd to give his utterance of (S3) the accolade of "true" or "correct," or to congratulate him for perspicacity in uttering it, even though his intention will be fulfilled. What is true or correct is his assessment of the circumstances—not his commitment to (S3). For (S3) it is a matter of the facts matching up to the words, rather than the words matching up to the facts. We can sum this up by saying that expressions of intention lack *paradigmatic truth-value.*

Sometimes, of course, we may express an intention by saying "It is true that I am going to do what he says I am" or "It is false that I am going to resign." Here, however, the words "true" and "false" belong to the actual formulation of the intention, rather than evaluate it. Hence their inclusion does not give the expression which includes them paradigmatic truth-value.

Sometimes, too, we may refer in a proprietary way to someone's *true* intention. But his true intention is simply what he *really* intends, just as his true belief is what he really believes. The proprietary is quite distinct from the non-proprietary, viz., the paradigmatic, "true." Thus *his* true belief need not be true, i.e., *a* true belief, a belief expressible by a true sentence. It may in fact be false. And it seems quite unnatural to describe *his* true intention either as true or as false.

You might object that my expression of intention in S3 must be paradigmatically true or false, because of its undoubted consistency with your report of my success by "He did it." But does the comparable consistency of my performative "I bet" with your informative "He bets" similarly entitle the former to truth-value? Not if we are to believe Austin. When I say "I bet," I succeed or fail in betting rather than say something true or false, correct or incorrect. And even if we are saying something true or false, etc., by virtue of succeeding or failing, the truth or falsity can hardly be described as paradigmatic.

3. Unlike some other types of utterance which lack paradigmatic truth-value, expressions of intention are, first and foremost, expressions of something mental. No doubt commands like "You shall report here at 2.30 P.M.," promises like "I will return it tomorrow," and performatives like "I bet" or "I promise," have to express such mental attitudes as wishes or judgments to have any point. But unlike expressions which are nothing but expressions of attitude, such expressions are used in the type of illocutionary act which is definable in terms of its perlocutionary effect.[3] Thus a command is, roughly, the sort of expression of a wish which conveys by linguistic convention that, should the person addressed fail to take reasonable steps toward satisfying the wish, he will become liable to some penalty.

---

[2] Cf. Rankin, "Referential Identifiers," *American Philosophical Quarterly*, vol. 1 (1964), pp. 233–243.

[3] A formula used by John R. Searle in *Speech Acts* (Oxford, 1969), p. 71.

4. In spite of paradigmatic truth-value an intention is an anticipation and judgmental in character. In forming an intention you subject your attitude to the criteria of consistency. Unless schizoid, you cannot knowingly intend to perform incompatible actions, nor can you form a categorical and unqualified intention to do something while subscribing to an inductively formed, and firm, belief that the action will not occur (see 1). Likewise your intention will be held by others to be inappropriate, if they believe that what you intend will not come about. By contrast, wish-attitudes—though like intentions in some respects—are non-judgmental. You may know that the realization of one wish precludes satisfaction of another, and yet wholeheartedly have both. Likewise your wish is not reckoned inappropriate by others, just because they think it will not be realized.

5. As a verb, "anticipate" like "believe" can be used to ascribe a disposition. Someone may anticipate a rise in salary next year, in the sense that if you take up the matter with him he may say, and will believe, that his salary will then be raised, even though the thought does not often cross his mind. But "anticipate" can also be used to refer to the sort of occurrence in which an anticipatory disposition is *realized*, i.e., to the fact that it *occurs* to someone that something, e.g., a rise, is on the way.[4]

Such occurrents, however, are admittedly somewhat singular. Sometimes they are the result of some process of deliberation, or resolution of uncertainty. As such *results*, they come about quickly or slowly, easily or with difficulty, etc. Sometimes, however, they come without any process. Hence they do not seem to fall within the Rylean category of achievement as a correlative of process. Sometimes, too, however they come, they may persevere continuously, or intermittently over a period of time, if on first occurring they refer to an occasion sufficiently far ahead. On such occasions they seem as complete in any small segment of their duration as in the whole. To this extent they are like states. On the other hand, surely it may occur to one quite instantaneously, without consciousness of time-span, that something will occur just before the happening does occur. To that extent they are more like changes from one state to another.

I may intend now, for instance, to do something rather exacting at a later date. Since what I intend will require all my resolution, my intention may remain in the forefront of my mind continuously, or intermittently as when I waver or spend time in sleep, right up to the occasion for action. And for as long as it continues, it will be complete, as an intention, at any one time, or for any one time-span, as at, or for, any other. On the other hand, perseverance through a stretch of time does not seem essential to the occurrence. The formation of intention may occur just at the moment of action.

6. Anticipating, intending, and certain other kinds of mental occurrent are rather singular in yet another way. They have temporal duality. Thus anticipating refers to something which is due to take place later, and remembering what has happened refers to what has taken place earlier.

7. Likewise they have an act-content structure. Each distinct occurrent is characterizable by a judgmental content, though one may have the same content as another. Consequently, where the content has paradigmatic truth-value, the occurrent, quite straightforwardly, belongs to the sort of thing which we do rightly or wrongly, correctly or incorrectly, etc. Where a perceptual judgment is false, the perceiver has misperceived, etc. Even where the content lacks paradigmatic truth-value in the way that the content of intending does, his intending is something the agent does appropriately or inappropriately, depending upon whether or not it is based upon correct assessment of the situation. Now it is by virtue of being either right or wrong, etc., that mental occurrents of these kinds have the status of being acts, though, of course, in contrast to intentional action they are more like ordinary happenings.

8. Anticipations of intention are *indexically committed* to the first grammatical person or some equivalent. The judgmental acts of different people can have a common content in cases other than intention. This holds even where the content of one person's judgment refers solely to himself.

---

[4] The Rylean dichotomy "Either disposition or occurrent, but not both" would require that the sort of occurrent anticipation which realizes an anticipatory disposition cannot itself be a disposition. A requirement on that basis would be unacceptable to those who go along (correctly, I believe) with William P. Alston ("Dispositions and Occurrences," *Canadian Journal of Philosophy*, vol. 1, 1971, pp. 125–154) and D. M. Armstrong (*A Materialist Theory of Mind*, ch. 6., § 6, London, 1968) in rejecting the dichotomy. There are, however, other grounds for regarding the realizing and realized anticipations as not dispositions in the same sense. Some of these are derivable more or less directly from later clues.

Thus, "I am over eighty" and "He is over eighty," as uttered by different persons, potentially express judgments with the same content. Such contents, accordingly, are uncommitted to the first, rather than the third, person. But one person's intention cannot have quite the same content as any sort of judgmental act of another. As an expression of intention, S3 is not translatable into S4 "He is about to *X*" uttered by another in reference to the same action of the same agent. There are at least four indications of this:

(a) Normally S4 expresses simple anticipation, not intention.

(b) This anticipation is formed on a twofold basis, viz., that the agent intends to *X*, *and* that he has the opportunity, including the capacity, to *X*. The anticipation of intention, on the other hand, is formed solely on the basis that the opportunity, including the capacity, obtains.

(c) The anticipation lacks the paradigmatic truth-value which the other anticipation possesses.

(d) Less normally, S4 might be short-hand for "I am going to induce him to *X*." But then it expresses the intention of another agent to perform a different action which will result in the first agent *X*-ing.

To appreciate the full significance of these considerations, we can turn once again to our previous analogy in which all four are reflected.

(a)′ Normally, to say "He is betting" or "He will return tomorrow" is respectively neither to bet nor to promise.

(b)′ These things are said appropriately on a twofold basis, viz., that their subject has uttered performances like "I bet" or "I will return tomorrow," *and* that the circumstances are appropriate for betting or promising. The performatives themselves, on the other hand, are appropriately uttered solely on the basis that the circumstances are appropriate.

(c)′ The performances lack the paradigmatic truth-value which their third person counterparts possess.

(d)′ Less normally, these counterparts might be shorthand respectively for something like "I guarantee he will bet" and "I will see to it that he will return tomorrow." But then their utterances become quite different performative acts of another speaker.

Now all four of these considerations point to a parallel indexical commitment in performative utterances, whereby they are untranslatable into third person counterparts. Despite the analogy, however, there are important differences between expressions of intention and other kinds of speech-act such as performatives. We have indicated these in clues 3 and 4. The true basis for the analogy will become clearer in the clues which follow.

9. Anticipations of intention are to some extent self-fulfilling. When the agent has an opportunity to act in opposing ways, it is essential to, and, if he has the capacity, sufficient for, his acting in any one that he intends to do so then—except in three types of case.

(a) He may be instantaneously competent to act appropriately. Thus he may be instantaneously competent to punch his tormentor on the nose, or to give his name when asked. Here his actions tend to be intentional, but not intended. He comprehends the situation in terms of the appropriate reaction to it—*as to be handled in such-and-such a way*, e.g., as a nose to be punched, or a question to be answered by giving his name.[5] One might say that he does not distinguish expressly between the challenge and his impending response. Notice that this act-oriented type of comprehension is antecedent to the act. That is why I questioned the inference in my introduction that intentionality without intention consists of nothing which antedates the intentional act.

(b) When an agent has an opportunityl including a capacity, to act in any of severa, opposing ways, one of the actions would have been necessary, through a sort of inertia, and the rest impossible, had it not occurred to him that he might do something other than the former. To begin with more universal types of case, in most situations it is necessary that I continue breathing or blinking, unless it occurs to me to desist for a time, perhaps to attend more carefully to a faint sound, or indistinct object, or to test my endurance. Similarly, in the case of well-established patterns of habitual behavior, such as driving home at the end of the day, once set off, we necessarily carry them through, unless it occurs to us to modify them in some respect. But even where neither the operation of our autonomic nervous system nor of habit is at issue, the same principle holds. When I am seated, normally the situation is such, e.g., the chair is sturdy enough, that I continue seated *necessarily*, unless it occurs to me to stand up, or slither down, etc. Accordingly, intention cannot be essential to the performance of any action which we would perform through the kind of inertia illustrated in these three ways, where it does not occur to us that we might act otherwise.

---

[5] F. H. Bradley has made a similar point in *Collected Essays*, vol. 2, (Oxford, 1935), pp. 39, 504, and 505. Michael Polanyi discusses analogous phenomena at some length in chap. 1, of *The Tacit Dimension* (Garden City, N.J., 1960).

(c) Where in performing one action we necessarily perform, or will perform another, the intention to perform the second is not essential to its performance. It is enough that we intend to perform the first. Thus, if I were to discern a star in the sky, if and only if I were to stop blinking for a while, I might in fact discern the star without intending to, if I had intended simply to stop blinking and did stop in fact. This example, however, suggests a complication to the present exception. What I do in stopping blinking is more likely to occur to me as trying to discern a star, than simply as 'stopping blinking. In fact, it may not occur to me as stopping blinking at all. Hence, for reasons given in (b) above, the intention to discern a star would be essential to my discerning it by stopping blinking. The same complication, however, might not arise in other cases of doing one thing in, or by, doing, or having done, another.

10. No agent can be blind to the distinction between his intention and non-self-fulfilling types of anticipation. If an anticipation of mine expressible by "I shall not inhale while under water" does not present itself to me as distinct from a non-self-fulfilling anticipation, I am probably in for a nasty shock after I submerge.

11. The self-fulfillment of intention must be distinguished from that of other kinds of self-fulfilling anticipation. The performance of an intentional action is quite unlike, for instance, a patient's recovery from what would otherwise have been a fatal illness only through confidence in his doctor's assurance that he will recover. "I am going to recover" is in these circumstances the expression of a mere anticipation—not of an anticipation of intention. The patient has been taken in to some extent, and possibly his anticipation could not have withstood the insight that his belief in recovery is an essential condition of his recovery. An anticipation of intention, on the other hand, does not seem to be in any way diminished, where the agent has explicit insight into its self-fulfilling role.

12. Anticipations of intention are also unlike those of the self-faith-healer who has acquired insight into the self-fulfillment of his anticipation of recovery. There are five indications of the difference.

(a) The self-faith-healer's insight is epiphenomenal to the self-fulfillment of his anticipation. The therapeutic power of his anticipation, in other words, is not dependent upon his insight, provided he really does anticipate. Whether the anticipation is induced by himself, his doctor, or by something else, is quite immaterial. According to 10, on the other hand, the agent's insight into the self-fulfillment of intention is essential to the self-fulfillment.

(b) Success in administering faith-therapy to oneself seems dependent upon the degree of confidence, or certitude, with which one can sustain the anticipation. The self-fulfillment of intention, on the other hand, is not tied to certitude to the same extent. Sometimes the strongest anticipation of intention an agent can commit himself to is expressible by nothing stronger than a formula like "I shall try to X." The uncertainty about the possibility of X-ing which this manifests need not interfere with successful performance. True, a person who doubts his ability to pronounce unslurred a tongue-twister like "The Leith police dismisseth us" may have less success than someone more confident, even though both suffer from the same degree of inebriation. On the other hand, his uncertainty may have the contrary effect, by making him exercise greater care.

(c) The ability of the self-faith-healer is an ability which, upon gaining insight into the therapeutic value of the anticipation of recovery, he has to acquire. Or at least he has to possess the pre-existing ability not to be put off by his insight. He has the ability despite the insight. On the other hand, there is surely no question of agents being put off intending upon acquiring insight into the self-fulfillment of intention.

(d) The ability of the self-faith-healer is something which he may acquire and apply for a purpose. He anticipates his recovery *in order to* recover. This anticipating may in fact be sufficiently deliberate to be intended as well as intentional. It may be intended as a means to an end, i.e., the end of recovery. But in that case the anticipating must be distinct from the intending to recover. It becomes part of the action intended, in the same way as reading aloud in order to amuse a child is part of the intended action of amusing the child. And if we were to ignore this conclusion by supposing the anticipating not to be distinct from the intending, we would be obliged to suppose that the intending was also an intentional action.

(e) Intending can be accompanied by insight into its self-fulfilling character without the anticipation incurring *on that count* the charge of irrationality. The self-faith-healer's anticipation of recovery does, however, contain an element of irrationality. A doctor, who has insight into the therapeutic effect that his assurance of recovery has upon his patient, can anticipate the patient's recovery in a rational way on the basis, among other things, that the patient has accepted his assurance. On the other hand, the self-faith-healer is not assured until he commits himself to the anticipation of his recovery. His anticipation consequently cannot have a complete rational basis. It may, of course, be rational in a pragmatic way, viz., as an effective means toward the recovery anticipated. But as shown in (d), intentions cannot be regarded as means toward their own fulfillment. Are we then to suppose that intending is irrational? Surely the agent who does what he does through the inertia described in 9(b) is less rational than the agent to whom it occurs that he can act otherwise, and who consequently does whatever he does intending to do it.

III. Why Intentions Are Not Causes

Sometimes in different ways these clues suggest a very simple interpretation, which can in turn be applied to the solution of our problem in Sect. IV. They indicate that intention is the sort of anticipation in which his anticipating is a *logically* necessary condition of the agent being about to act in the way he anticipates.

Clues 1, 3, and 4, on the one hand, and 9, on the other, bring the quasi-causal features of intention into sharper focus. Intentions anticipate what is intended not simply by way of coming before, but also in the way that cognitive acts or attitudes of anticipation anticipate. Likewise, with the exceptions noted, what is intended depends upon intention specifically as a necessary, not a sufficient condition.

If, however, the future occurrence of the anticipated action were *logically* independent of this necessary condition, then the agent's anticipating would presumably have paradigmatic truth-value. It would be "objective" in one of the senses commonly given to "objective," i.e., true or false logically irrespective of whether someone believes it to be true or false.[6] Hence clue 2, the lack of paradigmatic truth-value, indicates that this sort of anticipation is not logically independent of cognitive attitude.

If, on a contrary hypothesis, the anticipator's action were *causally* dependent upon his cognitive attitude, the logical status of his anticipation as paradigmatically true or false would not be diminished. If anticipations of intention are analogous to performatives, causally operative anticipations would be analogous to certain perlocutionary acts. Thus if you frighten yourself by saying "I frighten myself" what you say, unlike a performative, is paradigmatically true.

The indexical commitment of intention, to which clue 8 testifies, provides further evidence that the necessary condition is logical. Where an indexically *uncommitted* proposition about someone refers to him in the first person, the expression does not *state*, but only *indicates*, that the person expressing the proposition as his belief is the person to whom the proposition refers. At least a distinction between *stating* and *indicating* is needed, if we ever want to say that an utterance whereby one of us refers to himself, or herself, by means of "I" expresses the same proposition as an utterance whereby another of us refers to the same person by means of "you," "he," or "she." Where, on the other hand, an anticipation is indexically committed to the first person, what its expression *indicates* is also implicit, at least, in what it states. What the speaker states as his anticipation logically implies that he anticipates as stated. And what he so implies is exactly what his first person expression indicates. Consequently, his anticipating must be a logically necessary condition of his being about to act as stated.

This conclusion is reinforced by clues 10 and 11. Since the agent's anticipation of intention is essential to his being about to do what he anticipates, it is to that extent self-fulfilling. And since it is logically essential, the agent must have at least implicit knowledge of the difference between his anticipation of intention and any non-self-fulfilling anticipation.

Clue 12, in conjunction with 10 and 11, precludes the possibility that the agent's anticipating is

---

[6] In this sense the truth of a belief may fall short of objectivity in at least two ways. Either sincerity of belief is logically necessary or logically sufficient for the truth of the belief. Anticipations of intention exhibit only the former shortcoming. The belief that one is in pain is commonly regarded as exemplifying the latter.

causally a necessary condition of the anticipated action instead of, or in addition to, being logically essential. If the anticipating were causally a necessary condition, no insight, explicit or implicit, would be essential to its self-fulfilling operation. And, indeed, it is conceivable that any such insight might interfere with the self-fulfilling operation, as the case of faith-healing would seem to illustrate. Anticipations of intention, on the other hand, require implicit insight at least into their self-fulfilling operation, and are certainly not rendered inoperative by such insight.

If despite these arguments, however, they were causally self-fulfilling, they would have to resemble the anticipating of the self-faith-healer.[7] But they are quite unlike the latter in important respects, for reasons already given in clue 12. If the anticipating, on the other hand, is logically a necessary condition of its substance, none of these reasons apply any longer.

(a) The agent's insight cannot be epiphenomenal upon self-fulfillment of his anticipating, since to understand his anticipation he must understand, implicitly at least, that his anticipating is a logically necessary condition of his being about to act as he anticipates. Here, of course, I am arguing from the principle that to understand a proposition one must understand, implicitly at least, not only what is implicit in it, but also that it is implicit.

(b) The degree of certituce with which we hold a belief does not prevent what is implicit in it from being implicit in it. Hence if that we hold the belief is logically implicit in it, the absence of certitude does not prevent the belief from being logically self-fulfilling.

(c) Likewise we need acquire or possess no special ability to be able to withstand insight in the belief's self-fulfilling nature. The ability to withstand the insight is in fact inseparable from the ability to have the belief.

(d) Nor can the anticipating be regarded as something we undertake in order to fulfill, or as part of, the intended action, since logically it is not separate from the intending to perform the action.

(e) And finally, the anticipating cannot be regarded as irrational. In a logically true proposition like "If this is a rose, then it is a flower," the truth of the consequent is a logically necessary condition of the truth of the antecedent. But the truth of the consequent is not one of the truths on the basis of which we should infer the truth of the antecedent. Similarly, where the anticipating is a logically necessary condition of what we anticipate, there is no question of its presence being required as epistemic basis for itself.

It will be observed that the arguments in this section are of two main kinds. The first kind represents the connection between an anticipation of intention, and its substance as logical, more directly than the second. And the second represents the connection more directly as non-causal than the first. This double thrust protects my interpretation from the type of criticism to which the first type of argument has been prone in the past. Because causes are often described functionally in terms of their tendency to produce certain effects, the connection between them, when so described, and these effects is logical as well as causal. Hence, certain standard arguments, which establish a logical connection between intention and intended action, fail to show the connection to be likewise non-causal. For the latter purpose it is not enough to point out that, typically at least, intentions are specified in terms of the description of the action as intended, or that "other things being equal," it is logically necessary that given the intention, the action will occur. Now possibly my first kind of argument might turn out to be subject to this type of criticism. But the second kind does not seem vulnerable in the same way.

IV. How One Occurrence Can Be a Necessary Condition of the Subsequent Occurrence of Another Without Being a Causal Condition

A distinction between three forms of temporal series has now become overdue.

A. The series of stages in the "passage" of any one event from the future, through the present, into the past, whereby that event acquires perspective from successive moments of time.

B. The series of events which pass through these stages, stretching from the past through the present into the future, but not considered from the perspective of any particular moment of time.

---

[7] My method of refutation here does not assume that any mental act whatsoever, whether the anticipation of a self-faith-healer or anything else, ever is genuinely a cause. The assumption, however, would be hard to avoid were it not the case that anticipations of intention do not function as causes, and that other more cause-like mental acts play a constitutive role in intentions.

C. The events in the B-series when arrayed, from the perspective of any given moment of time, in a descending scale of pastness merging, through the present, into an ascending scale of futurity.

Thus the seasonal sequence of a bud appearing on a twig, a leaf sprouting, its growing, its withering, and its falling, is a segment of the B-series. Whether it is last, this, or next year's sequence does not affect the relative order in which these particular B-occurrents come. But, from the perspective of successive seasons, each of them is the basis of an A-occurrent belonging to an A-series. Thus, at any given instant in early spring the appearing of a bud that year is future, in later spring present, and in summer, autumn, and winter more and more remotely past. Finally, from the perspective of an instant in any one season, the whole B-series of seasonal changes in that year is the basis of a C-series. In summer the budding and sprouting are respectively more and less past, full growth is present, while withering and falling are respectively less and more future. Notwithstanding their membership of a C-series, however, these components are also A-occurrent. In fact the C-series peculiar to any point in time can itself be regarded as a complex A-occurrent belonging to a sort of orchestrated A-series which consists of the simultaneous changes in A-characteristics which affects every member of the B-series at successive moments of time.

Obviously some of the three series are not completely independent of others. For instance, ultimately[8] the subjects of the A-series and C-series qualifiers "being past," "being present," and "being future" are all B-occurrent. This has led to a radical difference of opinion, over the past fifty years or so, as to whether the A-series is, or is not, more fundamental than the B-series. I take the former alternative for granted here, though I have argued for it extensively elsewhere.[9]

The point of these preliminaries lies in the fact that the A-series can reasonably be represented as non-causal. No doubt it is universally true that every event that ever has happened, or will have happened, has earlier been or will have been about to happen—at least if the past is infinite. But, to put it mildly, a strong case can be made against supposing this to be an empirical generalization or law of nature, though explanations of the universal truth may radically differ. Both the B-series and the C-series, on the other hand, are causal in the sense that its components are at least *capable of* forming a causal chain, even though sometimes this sequence may not be a causal sequence. Thus, on our initial definition of "cause," our illustration of a B-sequence constitutes a causal chain, and any C-sequence, from which this B-sequence is an abstraction, must do so too. Since budding is causally prior to leafing, it being further in the past is a causal condition of leafing being nearer in the past.

In accepting on the strength of clue 5 that intentions are occurrent we must, accordingly, employ some circumspection. To which series does the succession of intention and action intended belong?

Clue 5 itself suggests the answer. Intentions can persist or be instantaneous, and where they persist they are complete in any temporal segment. A B-occurrent, on the other hand, can persist completely in all its segments, if and only if it is a state of something. And it can be instantaneous, if and only if it is a change from one state to another. But any specific type of B-occurrent must be exclusively a type of state, e.g., a temperature, or exclusively a type of change of state, e.g., a rise in temperature. Hence intention cannot be a type of B-occurrent.

Intending to X, on the other hand, is just like the A-occurrent to which (with the qualifications noted) it is logically essential, viz., *being about to X*. Being *about to X* is something *we are* before we X. We may *be* it over a period of time or intermittently, and in each case we may be about to as completely throughout a small temporal segment as throughout a large. But, in spite of all that, we may be about to quite instantaneously, right at the moment at which the action begins. Intending, then, is the same type of occurrent as that to which it is logically essential, viz. an A-occurrent.

---

[8] An A-occurrent based on a B-occurrent may, of course, itself be the basis of higher order A-occurrents and so on ad infinitum. Thus, at successive moments, the pastness of something now about to happen later will become less and less remotely future.

[9] Most recently in "Existence and Time" (*Philosophy and Phenomenological Research*, vol. 27, [1966] pp. 199–215), and "Past and Future: A Revision of Reichenbach's Indeterministic Distinction" (*The Philosophical Quarterly*, vol. 17 [1967], pp. 317–333). The thesis is (a) that the distinction which we draw by means of tenses and otherwise between past, present, and future has an ontological basis in a one-many indeterministic causal relation between events, (b) that in turn this indeterministic causal relation can only be properly conceived in terms of the distinction between past, present, and future. Both papers, however, require modification in the light of my present conclusions about mental occurrents such as intention. Such occurrents must be transferred from the category of causally-related occurrents into that of the causal relations which hold between the latter.

Clues 6 and 7 suggest this too. My being about to *X* has a temporal duality which complements that of anticipating *X*-ing. My *X*-ing has thereby a referred sort of occurrence. It is referred to the earlier moment at which I am about to do it. And the A-occurrent belonging to that earlier moment is characterizable in terms of my later *X*-ing, viz., as my being about to *X*.

Now this kind of referred occurrence can be understood as the matrix of the act-content structure of our intending, and of certain other mental acts which our intending involves. My *X*-ing characterizes not merely my being about to *X*, where I intend it, but, as we saw, it may characterize my intending as well. But my intending to *X*, except in the context of the opportunity, including my capacity to *X*, is only a logically necessary, not a sufficient, condition of my being about to *X*. Consequently, my *X*-ing characterizes my intending to *X* in a way which, unlike my being about to *X*, does not require the *X*-ing actually to take place.

Presumably here is the source of what is sometimes so confusingly, or perhaps pregnantly, described as the intentionality of certain belief-attitudes other than intention. What is believed, i.e., the content by which these attitudes are characterizable, may not be now, or thereafter, or ever have been, realized. As some say, these contents have intentional inexistence. And it is the susceptibility of contents to this spiritual malady that confers correctness or incorrectness upon the attitudes they characterize. It would occupy, however, much more time, space, and attention than we can spare here to show just how the act-content structure of the variety of mental acts depends upon their respective involvement in intention. Our immediate concern is to compare the occurrence of intending to *X* with that of being about to *X*.

Granted, then, that intending is A-occurrent, why is it yet so cause-like? A complete answer must fall into two stages. In the first place, the sequence consisting of my now intending to *X* and my now being about to *X* is extraordinarily like C-serial causal sequences. For instance, it is extraordinarily like the early spring C-serial causal sequence consisting of a bud now appearing and a leaf now being about to sprout. But it could only be C-serial if intending, whether earlier, now, or later, were B-occurrent.

The false likeness to C-serial causal sequences, however, only provides part of the answer. My now intending is an essential part and, under the appropriate circumstances, partly identical with my now being about to *X*. Hence, as well as not being in C-serial sequence, these two items are not in any temporal sequence at all. They are not distinct in the way that C-serial couplings are distinct. Should we then say that intending is not really as cause-like as I originally allowed, i.e., that intention does not precede what is intended? Obviously that won't do.

The rest of the answer, fortunately, is quite simple. My intending now does not precede my now being about to *X*. But it does precede my later having *X*-ed, in the same way as my now being about to *X*, of which my intending is an essential part, precedes my later having *X*-ed. In other words, it belongs to an A-series which has my *X*-ing as its sole B-occurrent basis. No doubt my intending also functions as a basis of an A-series. In other words, an intending may be past, present, or future. But it only so functions as a penultimate basis, i.e., in the way that my being about to *X* may itself be past, present, or future.

We have already remarked that some philosophers analyze the A-series in terms of the B-series. They commonly interpret our A-series ascriptions to any B-occurrent as either describing, or indicating pragmatically, the B-serial relations of that occurrent to the thought-tokens or cognitive acts which these ascriptions express. For instance, (S3) might be taken to mean, or indicate, that the speaker does *X* later than the thought-token expressed by (S3), where "does" is in a supposedly tenseless present.

Given the smog that currently envelops the status of mental acts, this, however, is an explanation of the less in terms of the more obscure. It rests upon the article of faith that mental acts, whatever else they are, must be B-occurrent. Hence, to rectify the thermal inversion, why not reverse the order of explanation, and classify them as A-occurrent instead?

V. Further Confirmation

Our two initial questions have each found an answer. Here, in conclusion, are a couple of clues to give these answers added point. They specify, in tandem, two of the most recalcitrant features commonly, or at least traditionally, attributed to mental acts.

13. Our mental acts are to some extent self-intimating. We have direct access to our own acts of awareness through acts of self-awareness. Consequently, if this is a universal feature of awareness, we should be aware of our self-awareness in infinite regression. On the other hand, though we can live up to such an estimate of our capacities for one or two steps in the regress, very soon thereafter our attention appears to flag.

Now, as noted in Sect. II, in our anticipations of intention we are implicitly aware of our anticipating. They are *implicitly* self-intimating. More specifically S5 "I am aware that I am about to X" is implicit in its subordinate clause, i.e., S3 "I am about to X." Accordingly, for at least one type of mental act, the apparent discrepancy between an infinite regression in self-awareness and our finite ability to regress presents no mystery. Where the whole of any sentence is implicit in a part, an infinite regress arises which we cannot hope to represent in more than a small part, viz., an *implicit* infinite regress. Our failure to attend to the infinitely regressive self-intimation is no more surprising than our failure to attend to the infinite number of evenly negated propositions implicit in any affirmative statement.

But what of other types of mental act? The most recalcitrant feature of self-intimation as an accomplishment of mental acts in general has been the apparent absence of any similar whole-in-part involvement in other types. How can the fact that I am now aware that there is a sheet of paper before me, and that a little earlier I went for a walk, be implicit in what I am aware of? However, there is no need for further involvements of this sort. Where we form an intention, it is in the light of the putative facts that we acknowledge in perception and memory, and of our thoughts about those facts. Certainly the facts are not self-intimating. But my perceiving, remembering, and reflecting, as constitutive of my intending, must be intimated in the self-intimation of the latter.

14. Mental acts are essentially private. Part of the intention of this claim has commonly been to distinguish their inaccessibility to more than one person from the sort of inaccessibility which can be overcome by exposing neurophysiological brain processes to direct or indirect sense perception. When, however, the claim is made on behalf of mind-body dualism, it is with the assumption that mental acts are B-occurrent. And, consequently, the intention becomes difficult to carry out. Secrets are essentially secret, and mysteries essentially mysterious. But in spite of that, the former are sometimes revealed and the latter solved. Similarly, why should the essentially private not become open to direct or indirect sense-perception?

If, on the other hand, mental acts are A-occurrent, the answer lies in the obverse fact that the objects of direct sense-perception are not A-occurrent, and that consequently this is true of the relevant kind of indirect sense-perception as well. By sense-perception, for instance, we may detect directly that a leaf is now falling from, as distinct from growing or withering on, a bough. We do not detect directly that the leaf is falling in the present as distinct from that it has already fallen in the past, or that it is going to fall in the future. We might *have detected* its falling if it had already fallen, and we might *be going to detect* its falling, if it is going to fall. But that is quite another matter. Discriminations between past, present, and future are not direct discriminations of sense. On the contrary, if my preceding analysis is correct, they belong to the self-intimation of intention, and the other mental acts constitutive of intention, instead. Through my intending I distinguish what I am about to do as in the future from what I have done as in the past, and my present from my past activities of perceiving, remembering, and generally thinking.

Nor do A-discriminations belong to indirect sense-perception, where the latter consists of inferences from directly perceived B-occurrents to the B-qualifications of those B-occurrents which we do not directly perceive. An example would be the perception of certain neuro-physiological processes in my brain, on the basis of directly perceived instrumental readings as interpreted by a neuro-physiological theory. No doubt this indirectness also depends upon our ability to make A-discriminations. Otherwise, we would be incapable of devising, confirming, and understanding any kind of interpretative theory. But, in its interpretative role, such a theory simply translates macrophysical into microphysical B-occurrent reports.

We can, of course, acquire knowledge of other minds[10] by indirect sense-perception as well, from what we directly perceive of how other people behave. But our discovery of what happened in the past, or of what to expect in the future, on the basis of what we directly perceive now is similarly indirect. Admittedly, then, the objects of indirect sense-perception need not be confined to B-occurrents. But where it extends beyond, it is further dependent upon the direct discriminations between A-qualifications which are to be found in the self-intimation of our mental acts.[11]

---

[10] That we do acquire this knowledge is readily explained as follows. Suppose quite generally that psychological verbs are closely related in status to the tense-inflection of verbs. Suppose, too, that the primary function of tense-inflection is to specify the kind of causal modality peculiar to branching systems of causally alternative happenings (see Rankin *op. cit.*). Our basis

University of Victoria
British Columbia

Received December 16, 1971

---

for ascribing psychological states to others will then be of the same probabilistic and inductive general sort that we have for making causal modal claims. Hence, neither the Scylla of analogical, nor the Charybdis of deductive inference from outer criteria to inner state need be involved as avenues of intimacy, except by way of approximation through myth.

[11] The analysis of intention in terms of a distinction between A and B- occurrents disposes of certain defects in my earlier account of intention in *Choice and Chance: A Libertarian Analysis* (Oxford, 1961). I acquired some of my dissatisfaction with the latter from R. L. Franklin's review, *Australasian Journal of Philosophy* (May, 1962), pp. 97–104 and from my former colleague, Max Deutscher. The present paper has also benefited from the criticisms, hints, and advice of, among others, Jonathan Bennett, Sam Coval, Danny Daniels, Terence Penelhum, Kirke Wolfe, and this journal's referee.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 1972

II. ON WEIGHTING CAUSES

RAYMOND MARTIN

CAUSAL explanations of particular results sometimes state, or imply, that a single factor was *the cause* of some result. Often, however, two or more causes are mentioned, one of which is said to have been *a more important* cause of some *particular* result than was another. The latter sort of explanation, which I shall call a "comparative causal explanation" (CCE) is a familiar and important feature of common sense and historical contexts of inquiry. For example, a student might claim that his not reading a specific text was a more important cause of his failing an examination than was his missing several lectures. Or, an historian might claim that Northern demand for tariff was a more important cause of the American Civil War than was agitation against slavery, or that Lenin's participation in the Russian October Revolution was a more important cause of its success than was Stalin's participation. The present remarks concern just the *weighting aspect* of CCEs. In what follows I shall offer partial analyses of what I take to be two of the most frequently employed descriptive senses in which the causes of particular results are weighted.[1]

I

Some philosophers have been sceptical of the weighting aspects of CCEs, and have even suggested that the causes of particular results cannot be "meaningfully" or "objectively" weighted. J. S. Mill's insistence that no "condition" of a result has a "closer relation" to that result than another, since each is ' equally indispensible to the production of the consequent" ([9], p. 214) is both a symptom of this suspicion and an indication of its source. Like Mill, some philosophers have claimed that the causes of some particular result must each be necessary for that result in some sense of "necessary" that does not admit of degree, and seem to have assumed that one cause of some result was more important than another only if it was "more necessary" (in the same sense of "necessary") for that result than the other.[2] Hence, no cause of some result was more important than another. However, no reason has been given for accepting the assumption which leads to this conclusion.

While some philosophers have had their doubts about the propriety of weighting the causes of *particular* results, most would readily concede that it is proper to weight the causes of *types* of results. For example, it is clearly meaningful, whether or not it is true, to claim that driver negligence has been a more important cause of highway automobile accidents than has mechanical failure in automobiles. Although there might be disagreement about exactly how the weighting aspect of this claim should be understood, it is plausible to suppose that in many contexts it would simply be the claim that driver negligence has been a cause of automobile accidents more frequently than mechanical failure has been a cause of automobile accidents.

Among those philosophers who have claimed that the causes of *particular* results *can* be meaningfully weighted, there has been an unfortunate tendency to assimilate the problems of accounting for the weighting aspect of CCEs to the problems of accounting for the weighting aspect of causal explanations of *types* of results. Those who have taken this approach, which includes almost every philosopher who has proposed an analysis of the weighting aspect of CCEs, tend to agree that the *weighting aspect* of most CCEs implies a frequency claim, i.e., a claim about the relative frequency of occurrence of factors of the same types as the causes weighted and the result explained. For example, Sidney Hook once claimed that a cause of a particular result was more important than other causes of that result "when events of this type take place more frequently than events of other types in relation to the type of occurrence under investigation" ([3], p. 114). And Ernest Nagel has characterized the following sense of "more important":

---

[1] There are also non-descriptive or "normative," senses in which the causes of particular results are weighted. See, for example, [4], p. 44.

[2] Consider, for example, J. L. Mackie's remark that "they may be distinguished as predisposing causes, triggering causes, and so on, but it is quite arbitrary to pick out as 'main' and 'secondary' different moments which are equally non-redundant items in a minimal sufficient condition, . . . ", [6], p. 253.

> Suppose that the joint presence of *A* and *B* is not a necessary condition for the occurrence of *C*, but that *C* occurs either when *A* is present conjointly with *X*, or when *B* is present conjointly with *Y*, where *X* and *Y* are otherwise unspecified determining factors; and suppose also that *A* in conjunction with *X* occurs much more frequently than does *B* in conjunction with *Y*. In that case, *A* might again be said to be a more important determinant of *C* than is *B*. Thus, suppose that automobile accidents happen either because of the negligence of motorists or because of mechanical failures in relevant parts in automobiles, and also that the frequency with which such mechanical failures cause accidents is very much less than the frequency with which negligence is responsible for them. In that case, negligence would be a more important cause of automobile accidents (in this third sense of the phrase) than would mechanical failures. Using analogous assumptions, a historian might similarly conclude that Austria and Germany's fear of Pan-Slavism was a more fundamental reason for the outbreak of World War I in 1914 than was the assassination of the Archduke Franz Ferdinand at Sarajevo. ([12], pp. 585–586; cp. [11]).

Both Hook and Nagel are claiming, as have others (e.g., [1], pp. 191–192; [2], p. 215), that there is a relatively familiar sense of "more important" such that (roughly) it may be said of particular factors, *A* and *B*, which were each a cause of a particular factor, *P*, that *A* was a more important cause of *P* than was *B*, if the frequency with which a factor of type-*A* has been a cause of a factor of type-*P* is greater than the frequency with which a factor of type-*B* has been a cause of a factor of type-*P*.

But this proposal must be rejected. It would be curious, if not downright wrong, to claim that driver negligence was a more important cause of some *particular* accident than was mechanical failure, merely because both were a cause of the accident in question and the frequency with which driver negligence has been a cause of automobile accidents is greater than the frequency with which mechanical failure has been a cause of automobile accidents. If there were such a sense of "more important," one could weight the causes of a particular automobile accident, once he knew what they were, without knowing any of the other details of the accident. Or, one could claim of an accident of whose details he was completely ignorant, "One thing is certain, if driver negligence and mechanical failure were among the causes of *this* accident, then driver negligence was a more important cause of *this* accident than was mechanical failure." Or, suppose there were just ten cities, six of which suffered from air pollution caused by industrial emissions, but not from air pollution caused by automobile emissions, three of which suffered from air pollution caused by automobile emissions, but not from air pollution caused by industrial emissions, and one of which, the tenth city, suffered from air pollution caused both by industrial emissions and automobile emissions. Surely there is no familiar sense of "more important" in which it would *follow* from these suppositions that industrial emissions were a more important cause of air pollution in the tenth city than were automobile emissions. These examples suggest that the weighting aspect of many CCEs may not imply a frequency claim. Hence, a fresh look at the weighting aspect of CCEs is in order.

II

Consider the following claims: I've gained so much weight since the last time I saw you partly because I'm getting less exercise now than before, but mostly because I'm eating much more. The coffee is bitter today, when normally it is not, partly because the pot was not clean, but more because the grounds had been used once before. Examples such as these are a familiar and relatively unproblematic feature of common sense contexts of inquiry. Yet the weighting aspect of these simple examples may be understood in terms of a model that is also appropriate to a sort of CCE that is frequently found in historical studies.

The weighting aspect of the explanations in the examples above conforms to the following model:[3]

(D1) *A* was a more important cause of *P* than was *B*, *if*
- (1) *A* and *B* were each a necessary cause of *P*, and
- (2) there is some appropriate state, ϕ, such that:
  - (i) *A*, *B*, and *P* are each deviations from ϕ, and
  - (ii) on the occasion in question, had *A* occurred as it did and *B* occurred as in

---

[3] Compare (D1) with Nagel's second sense of "more important," [12], pp. 584–585, item 'b'.

> ϕ, a result would have occurred more similar to P than had B occurred as it did and A occurred as in ϕ.

(1) of (D1) requires clarification in at least three respects. First, to say that A and B were each a *necessary* cause of P is to say that A and B were each a cause of P, and *on the occasion in question* no sufficient cause of P was present which was causally independent of either A or B; it is *not* to say that a factor of type-A and a factor of type-B are each necessary for the occurrence of a factor of type-P. Thus, someone's discarding a lighted cigarette may have been a necessary cause of the occurrence of a particular forest fire even though forest fires are not always preceded by a factor of this type. (I have discussed this sense of "necessary" in more detail in [8].) Second, ordinarily, when one says that one factor was a more important cause of some result than was another factor each factor was a necessary cause of the result. While this fact is reflected by (D1), it is not essential to the truth of a CCE. One may be unsure, and hence not wish to claim, that some factor was a necessary cause of some result, and still wish to assert that, at any rate, another factor was a more important cause of that result than the one in question. For example, one might claim that economic factors were a more important cause of some political decision than were ideological considerations, without also claiming that ideological considerations were a necessary cause of the decision. Or, one might claim that unconscious needs were a more important cause of some particular aspect of behavior than were conscious reasons, without also claiming that conscious reasons were a necessary cause of that aspect of behavior. However, in what follows we shall avoid a certain amount of circumlocution by restricting ourselves to examples in which each of the factors weighted was a necessary cause of the result. This should not be misleading, particularly if one remembers that D1 specifies only a logically sufficient condition for one factor's having been a more important cause of some result than was another. Thirdly, I shall assume that neither A nor B was, in the context of inquiry in which the explanation in question was provided, a "mere condition" for the result explained, and hence inappropriately regarded either as "a cause" or "the cause" of that result. For example, in most common sense contexts of inquiry, the normal presence of oxygen would not be regarded appropriately as either "a cause" or "the cause" of someone's weight gain, although we might insist that "strictly speaking" (i.e., independent of contextual considerations) the normal presence of oxygen was a cause of that person's weight gain. (I have considered this aspect of explanations in terms of "the cause" in [7]; analogous considerations are applicable to CCEs.)

(2) of (D1) requires a more extended discussion. Consider the first example above. If (D1) is to explicate the sense in which the individual's eating much more was a more important cause of his weight gain than was his getting less exercise, there must be some appropriate state such that his eating more, his getting less exercise, and his weight gain may each be regarded as deviations from that state. In general, identifying this actual or hypothetical state is often an easy matter in the context in which a CCE is offered, at least if the weighting aspect of the CCE is unambiguous. In our example, the person's eating more is a deviation from a prior level of food consumption, and his exercising less is a deviation from a prior level of exercise. The result, his weight gain, is a deviation from a prior weight level. When causes are weighted in the sense explicated by (D1), the weighting aspect of the explanation will not be clear unless it is clear from the context in which the explanation is provided which is the *appropriate* actual or hypothetical state from which A, B, and P are to be regarded as deviations.

Clarifying (2)(ii) of (D1) is more problematic, for we must clarify what it means to say that "a result would have occurred more similar to P." Here, we must recognize an important difference in results which are explained. First, consider cases in which it would be significant to say that had conditions been different, more, less, or the same amount of the result explained, P, would have occurred. For example, the result explained may be someone's weight gain or the bitterness of some coffee; and had conditions been different, a person might have gained more, less, or the same amount of weight, and the coffee might have been more, less, or equally bitter. When it is significant to say that had conditions been different, more, less, or the same amount of P would have occurred, then (2)(ii) of (D1) is true *if and only if* it is true that on the occasion in question, had A occurred as it did and B occurred as in ϕ, more of P would have occurred than had B occurred as it did and A occurred as in ϕ. Thus, the CCE in our first example is true if the individual's eating more and getting less exercise were each necessary causes of his weight gain, and on the occasion in question had he eaten as he did and exercised at his prior level, he would have gained more weight (i.e., more of his weight gain would have occurred) than had he exercised as he did and eaten at his prior level.

Secondly, consider cases in which it would *not* be significant to say that had conditions been different, more, less, or the same amount of the result explained would have occurred. The result explained may be someone's failing an examination for which the only grades were "Pass" and "Fail," or the result explained may be someone's death; and there is no *straightforward* sense in which it would be significant to say that had conditions been different, a person might have failed the examination more, less, or the same amount, or had conditions been different, a person might have died more, less, or the same amount. When it is *not* significant to say that had conditions been different, more, less, or the same amount of P would have occurred, we shall stipulate that (2)(ii) of (D1) is true *if and only if* there was some factor, P', suitably related to P, such that on the occasion in question, had A occurred as it did and B occurred as in ϕ, more of P' would have occurred than had B occurred as it did and A occurred as in ϕ.

Suppose that a student fails an examination for which the only grades were "Pass" and "Fail"; and the student claims, correctly, that he failed partly because he missed so many lectures, but more because he neglected reading one of the texts. The weighting aspect of his explanation will be correct, in the sense of "more important" under discussion, only if there was something, P', say, the value of the examination performance in the judgment of the grader, suitably related to P, such that on the occasion in question, had the student missed the number of lectures that he missed but read the text in question, the value of his examination in the judgment of the grader would have been greater than had he attended more lectures and not read the text in question. It is *in this sense* that had the student missed the number of lectures that he missed, but read the text in question, *a result more similar to* his failing would have occurred, than had he attended more lectures and not read the text in question. Similarly, in the case of explaining someone's death, the something, P', suitably related to his death, might be a deterioration (of the sort that led to death) of that individual's health. When causes are weighted in the sense explicated by (D1) and it is not significant to say that had conditions been different, more, less, or the same amount of P would have occurred, the weighting aspect of the explanation will not be clear unless it is clear from the context in which the explanation is provided which is the something, P', suitably related to P, in virtue of which the causes may be significantly weighted.

Our analysis is obviously incomplete in the absence of an analysis of "P' is suitably related to P." And I know of no way to analyze this expression. The best I can do is note two points. First, P' is suitably related to P only if P' was a cause of P. In the case of our examples, the grader's evaluation of the examination performance was a cause of the student's receiving a grade of "Fail," and the relevant deterioration of the person's health was a cause of his death. Secondly, examples of the sort under discussion are often such that the result to be explained is either the termination, or end state, of a process, or a product, or something akin to a verdict; and the something, P', suitably related to P, is generally some aspect of the process which led to the result, an aspect which admits of variation such that the more of P', all else being equal, the more likely that P would occur. For example, in the case of explaining someone's death. if the something, P', which was suitably related to his death, was a specific deterioration of that individual's health, then it is plausible to suppose that the more his health deteriorated in the respect in question, all else being equal, the more likely that he would die during the period in question.

Although we have conducted our discussion so far in terms of rather artificial, common sense examples of CCEs in which the causes are weighted in the sense of "more important" under discussion, more complicated examples of such explanations are a common feature of historical studies. For example, according to the historian Samuel Morison, the "main cause" of the slow recovery of the American economy after the Revolutionary War "was the double adjustment that the country had to make—from a war to a peace economy, and from a favored position within the British Empire to an independent status in a competitive world. A secondary cause of slow recovery was the right of each state to set up its own customs service" ([10], p. 282). Or consider Morison's claim that the knowledge that Washington and Jefferson were in favor of the new American Constitution was a more important cause of what public acceptance of that document there was prior to ratification than were the published essays of Madison, Hamilton, and Jay ([10], p. 314). The weighting aspect of these explanations may be understood, straightforwardly, in terms of the analysis expressed in (D1). For example, the weighting aspect of Morison's second explanation above would be correct if both of the factors he cites were necessary causes of the level of public acceptance of the new Constitution that existed prior to its ratification, and if, under the circumstances present during this period, there would have been more public acceptance of that document had Madison, Hamilton, and Jay not published their essays supporting it, but the public known that Washington and Jefferson were in favor of it, than had Madison, Hamilton, and Jay published their essays supporting it, but the public *not* known that Washington and Jefferson were in favor of it.

III

Perhaps the weighting distinction among causes that is most frequently encountered in historical studies is that sense in which the more important cause is often (but not always) a relatively permanent feature of the situation under investigation, and it is supposed that given the more important cause and the "prevailing circumstances," it was likely that had the less important cause not occurred, a "substitute cause" would soon have been present, and a result similar to the actual result would soon have occurred. For example, an historian might claim that the Northern demand for tariff was a more important cause of the American Civil War than was agitation against slavery, on the grounds that the tariff issue was not going to go away, and while it was present "a war of more or less this kind would have broken out sooner or later, whether the slavery issue was there or not, whereas the slavery issue by itself would never have led to war at all" ([1], p. 192). Or, consider the disagreement among some historians regarding the importance of the assassination of the Archduke Ferdinand as a cause of the outbreak of World War I. Those who deemphasize its importance vis-a-vis other factors sometimes suppose that these other factors were relatively permanent features of the situation under investigation, and that given these other factors and the "prevailing circumstances," it was likely that had the assassination not occurred, something else would have soon triggered the outbreak of hostilities.[4] Historians who assign greater causal importance to the assassination tend to deny this supposition. (See [13], pp. 76–77.)

There are two proposed analyses of the present sense of "more important" that are worth considering. While different from each other, both are initially plausible. Moreover, both are mistaken. Understanding why they are mistaken will help in formulating a more adequate proposal. The first, Nagel's, is another example of an analysis which asserts the logical relevance of a frequency claim to the weighting of the causes of particular results.

> Suppose next that *A* is a contingently necessary condition for *C*, and that, although *B* is not, it nevertheless belongs to a set *K* of mutually independent factors (B, B_1, . . . , B_n) such that the presence of some one member of *K* is a necessary condition for *C*. Suppose also that the various members of *K* occur with about the same frequency, the frequency of occurrence being in each case considerably greater than the frequency with which *A* is present. Accordingly, since the frequency with which some member of *K* occurs is greater than the frequency of *B*'s occurrence, then, even if *B* should not be present when *A* is present, the necessary conditions for *C* may nonetheless be realized because of the presence of some other member of *K*. These stipulations specify a sense of "more important" that is perhaps most frequently intended when *A* is said to be more important than *B*. . . . It is perhaps in the present sense that Germany's adoption of unrestricted submarine warfare is alleged to be the "main cause" of America's entrance into World War I. ([12], p. 585.)

The letters, '*A*', '*B*', . . ., in Nagel's analysans must be variables ranging over particular factors of various specified types, rather than indices for specific particular factors. If they were the latter, the relative frequency of occurrence of each would be either 0 or 1, and the analysis would be obviously defective. Thus, we must suppose that the letters are variables ranging over particular factors of various specified types, and that the related frequency claims are about the frequency of instances of factors of these specified types. But a problem arises immediately. For before we can test the analysis, we must know how to specify the relevant types, given an example of a CCE. And there are many examples of relatively clear CCEs, of the sort one finds in historical studies, for which it is not at all obvious how to specify the relevant types.[5] Consider a variation on Nagel's own example. Suppose an historian claims that Germany's adoption of unrestricted submarine warfare was a more important cause of America's entrance into World War I than was the disclosure of the "Zimmermann Note." What are the relevant types in virtue of which this CCE is correct and in accordance with the analysis Nagel has provided? For example, is the disclosure of the Zimmermann Note an instance of the type, "1917 proposal of German-Mexican alliance," or (simply) "proposal of German-Mexican alliance," or "anti-American secret German proposal," or . . . ? The possibilities are infinite. But it is not clear how we are to decide among them.

---

[4] Thus, Nagel's example above, the claim that Austria and Germany's fear of Pan-Slavism was a more important cause of the outbreak of World War I in 1914 than was the assassination of the Archduke Ferdinand, seems more plausibly regarded as an instance of the present sense of "more important" than of the sense it was intended to illustrate.

Or, suppose that the relevant types are specified, but that no factors of those types, other than those that are now being weighted, have in fact occurred. (This seems at least a logical possibility.) In this case, the relative frequency with which factors of each type are correlated with this type of result is 1, and hence one of the causes could not be said to be more important than another, in the sense of "more important" under discussion. Moreover, factors that could not be weighted correctly, at a given time, in the present sense of "more important," might be weighted correctly at a later time when more factors of the relevant types had occurred. But this suggestion is counterintuitive, to say the least. Surely the meaningfulness of the claim, say, that Germany's adoption of unrestricted submarine warfare was a more important cause of America's entrance into World War I than was the disclosure of the Zimmermann Note does not depend on whether other factors of the relevant types have *in fact* occurred.

Even if we ignore the problems just discussed, satisfaction of the requirements of Nagel's analysans does not insure that the causes have been weighted correctly. Suppose that the explanation in question is the claim that a drought, A^*, was a more important cause of a specified level of crop failure in a certain region, C^*, than was an insect invasion, B^*. A^* is an instance of the type, A (say, "drought"), B^* of the type, B (say, "insect invasion"), and C^* of the type, C (say, "specified level of crop failure"). Let B_1 be the type, "breakdown of irrigation system." Suppose that in this region a factor of type A and either a factor of type B or a factor of type B_1 (but not both) are necessary for a factor of type C; and that factors of type B and of type B_1 each occur with about the same frequency, and each occur much more frequently than factors of type A. Suppose further that in this region it is extremely rare for either factors of types A and B or for factors of types A and B_1 to be jointly present. (The occurrence of a drought motivates the people in this region to take special precautions against an irrigation system breakdown, and *almost* always kills great quantities of insects.) If, *on the occasion in question*, A^* and B^* were individually necessary and jointly sufficient for C^*, then Nagel's analysis would entitle us to claim that A^* was a more important cause of C^* than was B^*. But it seems gratuitous to weight the causes this way in the example as so far sketched. And if we embellish the example a bit it can seem downright counterintuitive to do so. For suppose that circumstances are also such that in the sense specified in (D1), B^* was a more important cause of C^* than was A^*, that is, the insect invasion destroyed more of the crops than did the drought. In this case, it would clearly be counterintuitive also to regard A^* as a more important cause of C^* than was B^* (in the present sense of "more important") *simply because* in the example as sketched these factors satisfy the requirements of Nagel's analysans.

I shall not pursue the task of trying to tighten up the analysans of Nagel's proposal. I am sceptical that any analysis of the weighting aspect of CCEs will be successful if a frequency claim is a necessary part of the analysans. Aside from the difficulties of determining relevant types, it is doubtful that such an analysans, if non-circular, can logically insure the absence of crucial respects in which the situation in which the causes and result occurred differ from situations in which factors of the same types occurred.

Sten Nilson has recently proposed an analysis which he claims captures the distinction between "dispositional" and "precipitating" factors, at least when the former were more important causes of the relevant result than were the latter.[6] Nilson's analysis applies to two sorts of cases. In the first, two particular factors, A and R, were on some occasion individually necessary and jointly sufficient for some result, E. Moreover, given A, it was probable that the result, E, would occur. In the second, two particular factors, A and Q, were on some occasion individually necessary and jointly sufficient for some result, non-E. Again, given A, it was probable that E would occur, and hence improbable, given A, that non-E would occur. Nilson claims that after the fact,

---

[5] There is a closely analogous problem for those who advocate a frequency theory of probability. See [5], pp. 20–21.

[6] Dispositional causes may not always be more important than precipitating causes. Compare the numerous examples in [14], *passim*.

> . . . when *either* E or non-E had occurred, importance can be ascribed as follows: In the first case, when E has occurred, A is considered an important, and R an unimportant causal factor. In the second case, when non-E has occurred, A is considered an unimportant and Q an important causal factor. In the first case A's importance will be reckoned greater relatively to that of R the greater was the probability of E's occurrence. In the second case, the smaller the previously existing probability that non-E *would* occur, the greater the causal importance of Q is considered to be when non-E actually *does* occur. ([13], p. 74.)

Unlike Nagel's analysans, Nilson's analysans does not include, at least not explicitly, any frequency claims.

But, consider an example similar to one we considered earlier. Suppose that during most seasons in a certain region there is both a drought and the irrigation system breaks down. When both happen, the crops fail; when either does not happen, the crops do not fail. And for any given season, given that there was a drought *or* that the irrigation system broke down, it was probable that there was a crop failure; and for any given season, the probability that there was a drought is greater than the probability that there was a crop failure. In such a case, Nilson is committed to saying that on some occasion when there was a drought *and* the irrigation system broke down, which together caused the crops to fail, then the drought was more important than the irrigation system breakdown as a cause of the crop failure *and* (in the same sense of "more important") the irrigation system breakdown was more important than the drought as a cause of the crop failure, *and* the degree to which the drought was a more important cause of the crop failure than was the irrigation system breakdown is greater than the degree to which the irrigation system breakdown was a more important cause of the crop failure than was the drought.

Moreover, the problems with Nilson's analysis do not stop here. Suppose that a million tickets were sold for some lottery, and that each had an equi-probable chance of being drawn for first prize. Suppose further that Jones buys ticket #542, A, and that ticket #542 is drawn for first prize, Q, and that on the occasion in question A and Q were individually necessary and jointly sufficient for Jones's winning first prize, E. On Nilson's proposal we would have to say that Q was a more important cause of E than was A; in fact, considerably more important since the probability of E, given A, was rather slight. But such a claim is counterintuitive. Jones's buying ticket #542 seems at least as important for Jones's winning first prize as ticket #542's being selected. Clearly then, Nilson's analysis must be rejected.

The main difficulty with Nilson's analysis is that he seems to have made two erroneous assumptions. The assumptions are strictly analogous, and mistaken for strictly analogous reasons. Nilson seems to have assumed that if two particular factors, A and Q, were individually necessary and jointly sufficient for some particular result, E, and given A, it was probable [improbable] that E would occur, then given Q, it was improbable [probable] that E would occur. Hence the truth of the antecedent of this conditional insures that A was a more [less] important cause of E than was Q. But, as shown in the first [second] counterexample above, it is a mistake to suppose that if the antecedent of this conditional is satisfied, then, given Q, it was improbable [probable] that E would occur.

IV

In what follows I shall specify a set of requirements the satisfaction of which is logically sufficient for one factor's having been a more important cause of some result than was another factor, in the sense of "more important" under discussion. This proposal, while not without some difficulties, at least avoids the counterexamples to the previous proposals we have considered. Perhaps it will serve as a prolegomenon to a fully adequate proposal.

Philosophers commonly assume that it is at least typically the case that if one factor, A, was a cause of another, P, then A was a non-redundant member of a set of factors, which set both occurred and was sufficient for P (see [8]). A particular factor, A, is a non-redundant member of a set of factors $(A, \ldots)$, for a particular factor, P, if and only if the set of factors ($A, \ldots$), of which A is a member, is sufficient for P, but this same set of factors, excluding A, is not sufficient for P. Thus, if a set of factors (A, X), is sufficient for P, but X by itself is not sufficient for P, then A is a non-redundant member of (A, X) for P; if, however, X by itself is sufficient for P, and X does not "include" the factor indicated by 'A', then A is a redundant member of (A, X) for P. In what follows I shall confine myself to those CCEs in which the causes cited, one of which was a more important cause of some result, P, than the other, are each non-redundant members of at least one set of factors (i.e., the *same* set of factors), which set (or sets) occurred and was (or were) sufficient for P.

Consider the following proposal:

(D2) A was a more important cause of P than was B if for any factor, X, such that, A, B, and X were each a cause of P, and were each a non-redundant member of the set of factors, (A, B, X), which set both occurred and was sufficient for P, the probability was greater, given A and X, than it was, given B and X, that a factor of type-P would occur during an appropriate temporal interval.

D2 should be understood such that the factor, X, is neither identical with nor "includes" either A or B. Moreover, as in D1, we shall assume that neither A nor B was, in the context of inquiry in which the explanation in question was provided, a "mere condition" for the result explained, P; however, no such assumption is made in the case of X. D2 is intended to express a sense of "more important" such that if A was an "underlying" or "dispositional," and B a "precipitating" cause of P, then A was a more important cause of P than was B (in the sense of "more important" in which dispositional causes are typically more important than precipitating causes), if and only if, A, B, and P satisfy the requirements in the analysans of D2.

We are going to have to rely on contextual clues to determine what is to count as a factor of type-P and what is to count as the appropriate temporal interval; and in actual cases of CCEs the boundaries of both are likely to be fuzzy. Consider a simple example. Suppose that someone has a strong and persistent desire to do something, is presented with an opportunity to do it, and does it. Moreover, his strong desire was a more important cause of his doing what he did than was his being presented with the opportunity, (at least partly) because it was probable both that his desire would have persisted and that he would have been presented soon with an opportunity to satisfy it. For example, suppose that Jones, Smith's employer, dislikes Smith intensely, and wants to fire him the first time Smith makes a sufficiently serious mistake. Suppose also that Smith makes a mistake of the requisite seriousness, Jones fires him, and both Jones's intense dislike for Smith, A, and Smith's making of the serious mistake, B, were causes of Jones's firing Smith, P. In this case we would say that A was a more important cause of P than was B if both A and B were non-redundant members of some set of factors (A, B, $\ldots$), for P, and for each such set of factors, given that A and $\ldots$ were present, it was more probable that the set would be "completed" than it was given that B and $\ldots$ were present. In our example, the set ($A, \ldots$), would have been completed (we may suppose), had Smith made a mistake of the requisite seriousness. And we could embellish the example such that the importance of A would increase vis-a-vis B, all else being equal, the shorter the temporal span in which it was probable (and the more probable it was) that Smith would make a mistake of the requisite seriousness. Moreover, in our example a factor of type-P might be, depending upon context, a firing of Smith by Jones, or perhaps simply a firing of Smith.

D2 may also be illustrated in terms of an example borrowed from Nilson ([13], p. 75). Suppose that a lighted cigarette is discarded into an open powder keg, the keg explodes, and both the discarding of the lighted cigarette and the positioning of the keg were causes of the explosion. So far, there is no reason to regard either factor as a more important cause of the explosion than the other. But if we also suppose that the keg was concealed in a place where explosives are almost never placed, but lighted cigarettes are frequently discarded, and hence that an explosion was more probable, given just the positioning of the open keg and the other factors (excluding the discarding of the lighted cigarette) which made up the sufficient set(s), than it was, given just the discarding of the lighted cigarette and the other factors (excluding the positioning of the open keg) which made up the same set(s), we would say that the positioning of the open keg was a more important cause of the explosion than was the discarding of the lighted cigarette.

Analyses of collective violence provide a plentiful source of CCEs the weighting aspects of which apparently conform to D2. Such social phenomena are often precipitated by an incident which was a less important cause of the result explained than were other underlying or dispositional factors. For example, the Detroit race riots of 1943 were precipitated by a racial incident on the Belle Isle Bridge on the evening of June 20. Some observers regard this incident, it would seem correctly, as a less important cause of the riots than the atmosphere of extreme tension which preceded the riots ([14], p. 249n). How are we to understand the weighting aspect of this claim? Presumably, the incident and the atmosphere of tension were each non-redundant members of at least one set of factors, which set both occurred and was sufficient for the riots; and the riots (or similar riots at about the same time) were more probable, given the atmosphere and the other factors (excluding the bridge incident) which made up the sufficient set(s), than they were, given the incident and the other factors (excluding the atmosphere) which made up the sufficient set(s). It is because this was so (if, indeed, it was so) that the atmosphere of extreme tension is correctly regarded as a more important cause of the riots than was the incident on the bridge. Or, suppose that an assassination takes place and a war breaks out, and that the assassination and certain underlying factors were each a cause of the outbreak of the war. It would seem that we could weight these causes correctly, vis-a-vis one another, if they satisfied the requirements in the analysans of D2. Moreover, this seems to reflect historical practice. Historians would, I believe, downplay the importance of the assassination, vis-a-vis the underlying factors, if the latter and certain conditions which preceded the outbreak of war rendered it more probable that the war (or a similar war at about the same time) would occur than did just the assassination and those same conditions. (Compare [13], pp. 76–77.)

This concludes our discussion of the two senses of "more important" that we set out to analyze. There are, of course, many respects in which our analyses, particularly D2, require clarification. And there are other senses of "more important" that require analysis. (Analyses have been proposed for some of these. See [12], pp. 582–7.) However, it is my impression that the two senses of "more important" that we have discussed in detail are the descriptive senses of "more important" most frequently encountered among those CCEs offered in common sense and historical contexts of inquiry. I hope to have done enough not only to clarify somewhat these two senses of "more important," but also to diminish the attraction of those mistaken approaches to the weighting aspects of CCEs, the consideration of which initiated our discussion.[7]

University of Maryland

Received February 3, 1972

References

[1] Gibson, Quentin, *The Logic of Social Inquiry* (London, 1960).

[2] Hart, H. L. A., and A. Honore, *Causation in the Law* (Oxford, 1959).

[3] Hook, Sidney, "Illustrations," *The Social Sciences Historical Study; A Report of the Committee on Historiography* (Social Science Research Council), Bulletin 64 (1954), pp. 110–115.

[4] Jensen, O. C., *The Nature of Legal Argument* (Oxford, 1957).

[5] Kyburg, Jr., H., *Probability and the Logic of Rational Belief* (Middletown, Conn., 1961).

[6] Mackie, J. L., "Causes and Conditions," *American Philosophical Quarterly*, vol. 2 (1965), pp. 245-264.

[7] Martin, Raymond, "Singular Causal Explanations," *Theory and Decision*, vol. 2 (1972), pp. 221–237.

[8] Martin Raymond, "The Sufficiency Thesis," *Philosophical Studies*, vol. 23 (1972), pp. 205-211.

[9] Mill, J. S., *A System of Logic*, 8th ed. (New York, 1896).

[10] Morrison, S. E., *The Oxford History of the American People* (New York, 1965).

[11] Nagel, Ernest, "Some Issues in the Logic of Historical Analysis," *Scientific Monthly*, vol. 74 (1952), pp. 162–169.

[12] Nagel, Ernest, *The Structure of Science* (New York, 1961).

[13] Nilson, Sten, "On the Logic of Historical Explanation," *Theoria*, vol. 2 (1970), pp. 65–81.

[14] Smelser, Neil, *Theory of Collective Behavior* (New York, 1963).

---

[7] Work on this paper was supported by a grant from Macalester College and a grant from the American Council of Learned Societies. I am grateful to James Lesher and to John Vollrath for helpful comments on an earlier draft.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 1972

III. FALLIBILISM AND RATIONALITY

JOHN KEKES

I

THE purpose of this paper is to raise doubts about certain philosophical arguments in favor of the assumption that science is the paradigm of rational inquiry. It will be argued that scientific inquiry, understood in a certain way, rests upon presuppositions which need to be rationally defended, and cannot, except question-beggingly, be scientifically defended. There is, therefore, a prior enterprise, the defense of presuppositions, upon whose success rests the rationality of scientific inquiry. If this enterprise is successful, it, and not science is the paradigm of rationality, while if it is unsuccessful, not only does scientific inquiry fail as a paradigm of rationality, but also, it fails to be rational. The philosophical position that gives rise to these doubts is fallibilism.

According to fallibilism human beings are doomed to error. Not even the most secure, the best supported belief is free from the possibility of mistake. Certainty is an illusion, proofs are deceptive, knowledge, understood as justified true belief, is an impossible ideal. Fallibilism challenges the epistemological orthodoxy that regards knowledge as the attainable product of a reliable process of reasoning; it denies the authority, the reliability of reasoning, and consequently declares knowledge, regardless of its source, to be unattainable.

Two forms of fallibilism need to be distinguished. Weak fallibilism is the view that it is logically possible that any conclusion arrived at by reasoning is mistaken, and that this possibility can never be excluded. Weak fallibilism allows, however, that even though reasoning is inevitably prone to error, some beliefs may be rational and ought, therefore, to be accepted. Strong fallibilism goes further and denies that any belief can have rational warrant. Weak fallibilism cautions against error, and advises that all beliefs be held tentatively; strong fallibilism denies that reasoning can provide grounds for the acceptance of any belief. Reasoning, on this view, is a blind alley, an illusion, a vain pretence.

Strong fallibilism is a name for the philosophical position for which thorough-going scepticism is another label. When Hume declares that "If reasoning be considered in an abstract view, it furnishes invincible arguments against itself,"[1] he is expressing the strong fallibilist strand of his thought.

Weak fallibilism does not occur by itself in epistemology. For if it is accepted that reasoning is untrustworthy and beliefs should be tentatively held, it is necessary to ask what warrant there is for one's beliefs. It may be argued that while reasoning is fallible, we can learn from our mistakes. The possibility of learning from them is the hallmark of rationality, and it is what makes the growth of knowledge, understood as the acquisition of rationally supportable beliefs, possible. The most notable champion of this position is Popper, and the name he has given it is critical rationalism.

Weak fallibilism, it will be argued, leads to strong fallibilism—a progression disastrous for the rational aspirations of man—and Popper has not managed to stem the tide of unreason. The main purpose here is to establish the negative conclusion that there is a problem about rationality that both needs to be and has not been solved. But it can be solved, and the end of the paper contains an indication of what the solution might be.

II

The initial case for weak fallibilism can be made by showing how it fares against its traditional opponent: the justificational view of rationality[2]. Rationality has been taken, traditionally, to consist in accepting all and only those beliefs that are justified, or at least justifiable. The justifiability of beliefs depends on the logical possibility of proving them true or probable. Beliefs are proven by reasoning, and only beliefs demonstrable by a reliable method of reasoning can be rationally held. Weak fallibilism, however, questions whether there is a reliable method of reasoning. Supporters of the justificational view of rationality are obliged, therefore, to argue for their reliance upon reason.

---

[1] David Hume, *Dialogues Concerning Natural Religion*, ed. by N. K. Smith, (London, 1947), p. 135.

[2] For a discussion and criticism of this view see K. R. Popper's "Oracular Philosophy and the Revolt Against Reason," ch. 24 *The Open Society and Its Enemies* 4th rev. ed. (Routledge, 1963), and "On the Sources of Knowledge and Ignorance," in *Conjectures and Refutations* (New York, 1968) 2nd ed.; see also W. W. Bartley's *The Retreat to Commitment*, (Knopf, 1962), especially chs. 4 and 5, and "Rationality versus the Theory of Rationality" in *The Critical Approach*, ed. by M. Bunge, (New York, 1964).

One possible argument is presented by dogmatic rationalism. According to this view, there are some standards of rationality and the reliability of any instance of reasoning depends on its conformity to the standard. There is disagreement about just what are the standards. Obvious candidates are the Cartesian clear and distinct ideas, the deductive ideal of entailment by self-evident axioms, the traditional empiricist requirement of logical construction out of "the given," and the early logical positivist standard of reducibility to protocol sentences.

No matter what the standard is there is a fatal objection against appealing to it. If rational justification consists in appealing to standards of rationality, then the standards themselves can never be rationally justified. For if the standard is judged rational by its own testimony, then the appeal is question-begging. If the standard is justified by an appeal to another standard, then the same question arises about the credentials of that other standard and so the appeal leads to a never ending regress of justificational attempts. Dogmatic rationalism thus is an easy prey to weak fallibilism which can point at the ultimate arbitrariness of any standard.

Another justificational attempt to counter weak fallibilism is analytic rationalism.[3] The rational justification of beliefs consists in appeal to some standard of rationality, but it is denied that the question of the rationality of the standards can arise. The reason given is that "rationality" is supposed to have meaning only with reference to some standard. Questioning the standard of rationality is like questioning whether the standard meter rod is really a meter long.

The virtue of this position is that it avoids the necessity of opting between the fatal alternatives of begging the question and of hopelessly pursuing the justification that does not stand in need of justification. Analytic rationalists clearly recognize that if the question of the rational justification of standards of rationality is allowed to be asked, then rationalism must surrender to weak fallibilism. They strive, therefore, to rule the question out of order.

Analytic rationalism, however, fails. Part of the problem is that the nature of the account of standards of rationality is unclear. If it is supposed to be a *description* of how arguments, as a matter of fact, are judged, then analytic rationalism is simply false. Rhetoric, propaganda, flattery and threat, guilt and greed play a considerably greater role in the judgment of arguments than does the judicious weighing of evidence. If, however, it is supposed that analytic rationalism offers a *prescription* of how arguments ought to be judged, then it is perfectly proper to demand some support for the recommended standards. The need for the satisfaction of the demand is made much the more acute, because alternative standards—utopian, religious, mystical, aesthetic, political—are presented in frightening abundance. Nor does it help to avoid weak fallibilism to define "rationality" as "satisfying such-and-such standard," because the whole problem is whether or not arguments should be judged favorably by their satisfaction of those standards that analytic rationalists are inclined to accept.

The weak fallibilist need not be disconcerted by the charge that he makes use of the standards he attacks. For the fallibilist position is that *all* standards are arbitrary and hence all beliefs lack rational justification. The logic of the fallibilist position is a *reductio ad absurdum:* it is to assume the standards of the justificational rationalist and show that they are arbitrary.

If the justificational view is the only theory of rationality, then weak fallibilism is unavoidable. The very general reason for this is that justification is taken to consist in appealing to standards of rationality and these are neither self-evident nor are they capable of being rationally justified. The great danger is that if weak fallibilism reveals that the justifications offered by rationalists are, in the last analysis, no less arbitrary than appeals to faith, revelation, instinct, taste, or whatnot, then it is difficult to see how strong fallibilism could be avoided. And strong fallibilism heralds the death of reason.[4]

---

[3] This is the view of A. J. Ayer in *The Problem of Knowledge* (Baltimore, 1956), see especially pp. 74–75 and of P. F. Strawson "The 'Justification' of Induction" in *Introduction to Logical Theory* (London, 1952).

[4] The problems raised in this section are discussed in greater detail in my "The Sceptical Challenge to Rationality," *Metaphilosophy*, vol. 2 (1971), pp. 121–136.

III

Popper's critical rationalism sides with the weak fallibilist objections to the justificational view of rationality. Popper refuses, however, to draw an anti-rationalist conclusion from this. Weak fallibilism does not lead to strong fallibilism. It can be reconciled with rationalism, provided the justificational view of rationality is replaced by the critical view.

Popper takes rationality to be characterized by the critical attitude, and the natural sciences as the field where the critical attitude is most typically exemplified. Hence for Popper the philosophy of science is the key to epistemology. If it is possible to reconstruct the logic of science, then, thereby, questions about rationality, truth, knowledge, become answerable.[5]

Consider the development of science. In the pre-scientific stage there are a large number of myths, practices, beliefs, and assumptions about reality. These form the background, the tradition against which problems occur. Science starts with problems. Problems reveal that there is a tension between various traditional assumptions, or that the traditional assumptions conflict with observation. The solution of the problem is demanded by the necessity of coping with the environment and this requires having an accurate picture of it.[6]

The problem calls for a conjecture. A conjecture is a guess about the nature of reality. The purpose of the conjecture is to bring the traditional assumptions into harmony with observation in such a way that what was previously problematic will become understandable. The conjecture gives a picture of reality, its role is to depict how realitymay be so that one's assumptions would fit it. The fact that a problem occurred previously shows that the old assumptions did not quite fit reality, that they need to be revised. The conjecture is a possible revision.[7]

Given the background and the problem there are, of course, very many conjectures possible. All successful conjectures will reconcile traditional assumptions with existing observations and will offer a solution by proposing a picture of a segment of reality. But not all successful conjectures will be rational. What makes a conjecture worthy of scientific attention is that it is criticizable or falsifiable.

A criticizable conjecture is a rational scientific theory; an unfalsifiable conjecture is pseudoscientific, metaphysical. The criticizability or falsifiability of a scientific theory depends on it being specifiable what would have to happen for the theory to be shown to be mistaken.[8]

Popper's break with verificationsism, that is, with the view that the rationality of a scientific theory depends on the evidence available for it, is closely connected with the asymmetry of verification and falsification. A counter-instance to a theory tells much more decisively against it than does a verifying instance tell for it. This feature follows from the nature of theories. If a theory is a conjecture composed in part of traditional assumptions and past observations, then any theory will have some verifying instances built into it due to the possibility of repeating those observations which it is supposed to explain. The uselessness of evaluating a scientific theory on the basis of successful verifications can be recognized if it is noticed that it is a simple matter to generate successful verifications from scientific theories that are known to be mistaken.[9]

Instead of looking for observations that would bear out the picture of a segment of reality formed by a scientific theory, one should look for observations that could not obtain if reality was as the theory depicted it. Scientific theories issue prohibitions: such and such could not happen. The testing of a scientific theory comes to cajoling reality into yielding up those features which, according to the theory, are not there. To test a theory is to try to refute it. If a theory withstands testing, if one tried hard to refute it and failed, then the theory is acceptable and can be reasonably held until either it is refuted by some more ingenious test, or some new problem occurs that the theory cannot handle.[10]

---

[5] See for instance *The Logic of Scientific Discovery* (abbr.: LSD), (Hutchinson, 1959), p. 16, and "A Realist View of Logic, Physics, and History," in *Physics, Logic and History*, ed. by W. Yourgrau and A. D. Breck, (Plenum Press, 1970), especially p. 5.

[6] See for instance Sect. VI in "Truth, Rationality, and the Growth of Knowledge," (abbr.: TRGK) in *Conjectures and Refutations* (abbr.: CR), (New York, 1968).

[7] See for instance Sect. 6 in "Three Views Concerning Human Knowledge," also Sect. VII in "Science: Conjectures and Refutations," both in CR.

[8] For a detailed account of falsifiability see ch. IV in LSD.

[9] For an account of Popper's rejection of the verifiability criterion see "The Demarcation between Science and Metaphysics," in CR.

[10] From Popper's rejection of verificationism there follows the rejection of probability as a criterion for the appraisal of cientific theories. For Popper's views on induction and probability see LSD, especially ch. X.

A theory that is known to be mistaken presents clear evidence about the nature of reality: it tells us that reality is not as the theory depicted it. Some theories resist refutation, at least temporarily, and these are accepted, but not of course proven. Theories cannot be known to be true with any certainty; at best, they can be known to have survived sophisticated critical attempts and have presented a solution to our problems. This is the reason why they should be accepted. There is, however, nothing final or certain about surviving theories. Their day of reckoning is merely put off until the problems they inevitably generate they can no longer solve.

Knowledge grows, but this must not be thought of as a process whereby increasingly larger number of truths are discovered. The growth of knowledge consists in making bolder and bolder conjectures; conjectures that are exposed to criticism, survive it, and lead to the solution of increasingly more difficult problems.[11]

The growth of knowledge is motivated not only by problems, but also by an ideal: truth. The purpose of theories is not just to solve problems, but to solve them in such a way that the theory gives a true picture of reality. Truth is the regulative ideal of the growth of knowledge; it is the ideal limit toward which theories tend. Truth, however, is unattainable; the ideal can be approximated, but it can never be achieved. That this is so follows from what has been said up to now. For to claim that a theory is true requires that it be provable, and to prove a theory requires the elimination of the possibility of error. Not only is it impossible to eliminate all sources of error, but also even if this were accomplished, nobody could know that it has been done.

Theories must be judged on the basis of their problem-solving capacity, their boldness of conjecture, and their survival of criticism. But it does not follow that if a theory passed these tests, then it is true. All that follows is that it is closer to being true, that it has greater verisimilitude than do theories that failed the tests. Verisimilitude is a relative concept; it must always be expressed as the property of a theory *vis-a-vis* another theory. We know what a true theory would be like: it would be a picture of a segment of reality such that the picture would correspond exactly to reality. But we also know that one can never know that a theory is true, all that can be known is that it is the best, boldest, most informative solution to a problem and that while it could clash with observation, it does not. We have all the reason to accept such a theory even though we do not and cannot know that it is true.[12]

If the quixotic quest for certainty is abandoned, if it is realized that truth is an unattainable ideal, then the failure to achieve it need not be counted as a failure of reason. The function of reasoning is to solve problems, to provide conjectures about the nature of reality, and to submit these conjectures to criticism. Weak fallibilism is correct in this: reason is fallible. The fallibility of reason, however, can be turned to good advantage by the method of conjectures and refutations; a method that is exemplified by scientific enquiry. Weak fallibilism, therefore, does not lead to strong fallibilism. Mistakes are unavoidable, but we can learn from them.[13]

IV

Attractive as critical rationalism is, it fails because it has not avoided strong fallibilism. The view that scientific enquiry is the paradigmatic rational activity and as such it is the remedy against scepticism is open to serious objections. The point of raising these objections here is not that the rationality of the scientific enterprise is in doubt, but rather that the assumption that scientific enquiry is rational needs also to be rationally supported and Popper's attempts to do so are unsuccessful. The alleged progress of science is not an obstacle in the way of strong fallibilism.

The distinction between the context of justification and the context of discovery is essential to the rational appraisal of scientific enquiry. The circumstances surrounding a man's formulation of a theory are judged to be irrelevant to the justification of the theory. The psychological, sociological, and fortuitous factors influencing a problem situation bear no relation to the rational appraisal of the solution of the problem. What leads a man to say something is independent of the truth or falsity of what he says.

Popper explicitly accepts this distinction.

> This question of how it happens that a new idea occurs to a man . . . may be of great interest to empirical psychology; but it is irrelevant to the

---

[11] For a detailed statement of Popper's views on refutation, growth of knowledge, and truth see TRGK.

[12] For a discussion of verisimilitude see Sect. XI–XIV in TRGK.

[13] A succint summary of Popper's views is to be found in the 1962 Preface to CR.

> logical analysis of scientific knowledge. This latter is concerned not with *questions of fact* . . . but only with questions of *justification or validity* . . . .The task of the logic of knowledge . . . consists solely in investigating the methods employed in those systematic tests to which every new idea must be subjected if it is to be seriously entertained.[14]

The rationality of scientific enquiry, or indeed of any enquiry, would be fatally undermined if questions of justification could not be answered in a satisfactory way. To acknowledge that this happens is tantamount to admitting that the enquiry cannot be rationally appraised; that this unavoidably happens is one of the main contentions of strong fallibilism.

One attack on the distinction between the contexts of discovery and justification is based on the idea that the standards by which theories are appraised are themselves theory-bound, part of the fundamental assumptions of the domain in which the theory occurs. Hence appraisal is either question-begging, or it must occur with reference to context-independent standards, and, to say the least, it is very difficult to see what such standards could be.

Suppose, for instance, that the standard by which a theory is judged to be rational is that it has been reached by following a certain method. Two methods that enjoy widespread acceptance among philosophers and practitioners of science are intersubjective testability and quantifiability. Goethe's theory of colors is unquantifiable, Piaget's studies of the development of individual children are intersubjectively untestable.

Intolerantly, it may be said that since these cases contradict accepted methods of scientific enquiry, they are irrational. Guided by a more liberal spirit, they could be declared merely unscientific, and the possibility of a favorable judgment of their rationality with reference to some non-scientific standard may be allowed. But this is a very peculiar way of dealing with counterexamples. Clearly, both Goethe and Piaget thought that they were doing science, and it is unlikely that they had been unacquainted with the scientific methods. They seem, however, to reject the methods that most scientists accept. Those and indeed any methods can thus be questioned, and consequently adherence to them requires justification. The justification cannot merely be that this is how scientists do things, because scientists are prepared to criticize practices that conflict with their own. Is this criticism rational, or is it just a flaunting of prejudices?

The argument in favor of accepting such methodological requirements as intersubjective testability and quantifiability is that with their aid scientific problems can be solved. It would not at all do of course to identify scientific problems as problems tractable by the application of scientific methods. For if this suggestion were accepted, the whole scientific enterprise would become self-guaranteeing, since problems not amenable to scientific treatment would not be considered, and *simpliciter*, would not be considered unsuccessfully, consequently science could not fail.

It might be suggested that problems do not come hallmarked as scientific; problems just occur and as it happens science is an excellent device for solving very many of them. But this will not do either, for problems do not just occur. There are, as it were, no problems in nature. Events, occurrences, processes appear problematic only against the background of assumptions which shape expectations. If the expectations urge scientific solution of the problems, then the employment of scientific methods is no doubt amply warranted. But perhaps a man who refuses to follow the accepted methods does so because his expectations have been shaped by different assumptions, and hence different things strike him as being problematic. To judge the adequacy of his solutions by the scientific yardstick is like commenting adversely on a pointillist painter for not using straight lines.

The reply might be that not all problems arise against assumptions. There are universal problems that stem from the physiological needs of human beings and from the nature of the environment which has to be forced to yield up the satisfaction of these needs. This is no doubt true, but its truth presents no argument in favor of those assumptions which form the foundation of science. For there are and have been many civilizations that solved these basic problems without the application of the scientific methods. The scientific solution is *a* solution, not *the* solution. To judge other, non-scientific, approaches by applying to them scientific standards is of course question-begging. A successful argument for science being the paradigm of rationality must be based on a demonstration that those assumptions upon which science is based are rationally preferable to other assumptions.

If a grand label is wanted, these assumptions might be called "metaphysical." At any rate, they express *a priori* beliefs about the nature of reality. They are the organizing principles of experience; they are the devices we employ to make sense out of data that continually bombard each human being. The assumptions are certainly not necessary, they are not supposed to be true by definition. At the same time, they are not empirical either, since they would not be refuted by experience that seems to conflict with them. Experience cannot be used to appraise these assumptions, for it is the role of these assumptions to appraise experience by them. Kant called them synthetic *a priori* principles, Collingwood referred to them as absolute presuppositions, and Popper, no doubt, would include them into the "background" against which problems arise. *A priori* assumptions may be ontological, such as "There is a spiritual element in reality," or "Everything is made of matter," or "There exists a life force"; or they may be dichotomies in terms of which experience should be classified, for instance, "cause and effect," "mental and physical," "real and apparent," "transcendental and sensible;" or the assumptions might be about how nature works and how best to understand it, for example, "Nature allows no exceptions," "Nature is orderly," "God's will be done," "God does not play dice," "There is purpose in nature," and so on.[15]

---

[14] LSD, p. 31.

The acceptance of scientific enquiry as a paradigm of rationality requires a demonstration of the assumptions upon which scientific enquiry rests. But such demonstration cannot come from within science, since it would already presuppose what requires to be proven, namely, the assumptions of science.

Naturally Popper is aware of these pitfalls and attempts to avoid them. What makes the scientific enterprise preferable to other attempts at problem-solving is the cumulative, progressive nature of science. "I assert," Popper writes, "that continued growth is essential to the rational and empirical character of scientific knowledge; that if science ceases to grow it must lose that character." And he goes on to say: "Science is one of the very few human activities—perhaps the only one—in which errors are systematically criticized and fairly often, in time corrected. This is why we can say that, in science, we often learn from our mistakes, and why we can speak clearly and sensibly about making progress there. In most fields of human endeavor there is change, but rarely progress . . . for almost every gain is balanced, or more than balanced by some loss. And in most fields we do not even know how to evaluate change."[16]

This view of science has been contested by Kuhn.[17] Central to Kuhn's thesis is the distinction between normal and revolutionary science. The distinction rests on the notion of paradigm. A paradigm performs three main functions. First, it is a scientific achievement that has come to be regarded as a model or ideal of scientific research by scientists in a given period. It is an achievement that opens up new avenues of research, new ways of tackling problems; it establishes the foundations of a new way of looking at reality. The work of Newton or Darwin, for instance, constitute paradigms in this sense. Another aspect of the paradigm is that it determines the type of questions that scientists ask. If certain observations suggest problems, it is because they are at odds with expectations based upon the paradigm. So the paradigm acts as a regulative ideal prompting some questions, judging others as irrelevant or uninteresting, prescribing methods of research that are likely to be successful. In its third role the paradigm is a textbook, or an actual piece of research, that is, a concrete scientific achievement that functions as an educational tool for the training of future scientists. It is a showcase of good science.[18]

Normal science refers to a period of scientific research when there is an agreed-upon paradigm; revolutionary science occurs when scientists for one reason or another become dissatisfied with the existing paradigm and it is being replaced by another. Normal science, as the adjective suggests, is what science largely is. The overwhelming majority of scientists spend all of their professional time pursuing normal science, that is, research more or less determined by a paradigm. Scientific revolutions, the replacement of one paradigm by another, are very rare. When they occur, they determine for generations to come what normal science is to be.

---

[15] The nature of these assumptions is discussed in some detail in my "Metaphysics and Rationality," forthcoming in *Idealistic Studies*.

[16] Both quotations are from Sect. I in TRGK.

[17] See T. S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, 1962), and "Logic of Discovery or Psychology of Research?" in *Criticism and the Growth of Knowledge* (abbr.: CGK), (Cambridge, 1970), ed. by I. Lakatos and A. Musgrave.

[18] This elucidation of the notion of paradigm is Margaret Masterman's; see her "The Nature of a Paradigm" in GCK.

Kuhn agrees with Popper that science starts with problems, but he disagrees about the nature of the problems. The paradigm, according to Kuhn, imposes something like a grid upon nature. The grid has large gaps and it is not clear how to fill them. The paradigm gives an outline of nature; the task of normal science is to complete the outline, to fill in the gaps, to transform the loose grid into a finely meshed one. Scientific problems, then, are the difficulties encountered in the way of completing the task of normal science. Scientists doing normal science, in a sense, already know what nature is like: it is as the paradigm depicted it. Their problem is the application of the big picture to the particular case. Normal scientists are given the general answers, the rules by which the answers can be reached, and an excellent example of reaching the answer—they are given all this by the paradigm—and their problem is to reconcile a particular observation with the dictates of the paradigm. Their activity, Kuhn suggests, is better described as puzzle-solving, rather than problem-solving. The solution of a problem of normal science is not the discovery of some new truth about nature, it is a successful reconciliation of an observation with the expectations created by the paradigm.

But what happens if the reconciliation is not achieved? Is it not then that there is a refutation of the paradigm? Do scientists have to admit then that nature has said "No!" to their paradigm? Not at all. If a scientist fails to reconcile an anomalous observation with the paradigm, there are two options open. It may be concluded that the scientist failed to solve the puzzle, or the paradigm may be rejected. Normal science is characterized by the adoption of the first option. Only if the paradigm is already suspect would a conflict between it and an observation be interpreted as a criticism of the paradigm.

The role of criticism is thus profoundly different for Popper and Kuhn. The task criticism accomplishes, according to Popper, is that of getting nearer to the truth, of increasing the verisimilitude of our theories by eliminating mistakes. Mistakes are built into theories, and it is thus theories that suffer as a result of criticism. On Kuhn's view criticism may well leave theories intact and reflect adversely only on the skill, intelligence, or ingenuity of the scientist. As a result, criticism does not bring one closer to truth, it merely spurs the normal scientist toward more ingenious attempts at reconciliation. Popper regards science as an unceasing attempt to overthrow the paradigm; Kuhn sees science as increasingly sophisticated attempts at defending the paradigm. Popper thinks that normal science is atypical, Kuhn thinks that revolutionary science is atypical. Neither denies that the phase the other regards as important occurs, they deny that the other has correctly gauged its importance.

If Kuhn's view of science is correct, then Popper's claim fails. Science, then, is not distinguished from other areas of human endeavor by its capacity for progress. Progress in science is filling in the gaps left by the paradigm. But such progress is possible in any field that has a paradigm. If an historian accepts, say, dialectical materialism, his paradigm is provided by Marx and progress consists in explaining more and more historical events in dialectical terms. If a jurist accepts the codification of social behavior as the legal ideal, then progress will be recognized as the extension of the legal system to increasingly greater areas of social life. Paradigms breed progress, and scientific progress has no special place.

The disagreement between Popper and Kuhn is a family quarrel. They disagree about the nature of science, but they do not disagree about science being the rational enterprise *par excellence*. Kuhn in reply to the charge that he defends irrationality in science writes:

> To describe the argument as a defence of irrationality in science seems to me not only absurd but vaguely obscene. I would describe it . . . as an attempt to show that existing theories of rationality are not quite right and that we must readjust or change them to explain why science works as it does. To suppose, instead, that we possess criteria of rationality which are independent of our understanding of the essentials of the scientific process is to open the door to cloud-cuckoo land.[19]

This is a pure expression of the assumption that needs defending. Popper and Kuhn agree in making it, and they are at one in failing to defend it. The point of the argument in this paper is that unless the assumption is defended strong fallibilism prevails. For strong fallibilists need not deny that there is progress in Kuhn's sense; they merely deny that the paradigm is rationally justifiable. They detect no substantive difference between completing a paradigm, scientific or otherwise, and completing a jigsaw puzzle.

Kuhn's endorsement of science as the touchstone of rationality is a semantic decision. It amounts to advocating that we should use "rational" to refer to *scientific* enquiry in its primary sense. But definitions, as we have seen in criticism of analytic rationalism, will not assuage the strong fallibilist. Rationality, for it to be worth anything, must have something to do with truth, where truth is understood as correspondence between statements or theories and the world. Popper endorses this conception of truth and rationality, Kuhn denies it. In characterizing the difference between himself and other philosophers of science, including Popper, Kuhn writes:

---

[19] T. S. Kuhn, "Reflections on my Critics," p. 264 in CGK.

> There is another step . . . which many philosophers of science wish to take and which I refuse. They wish, that is, to compare theories as representations of nature, as statements about "what is really out there" . . . I believe nothing of that sort can be found . . . . If I am right, then "truth" may, like "proof," be a term with only intra-theoretic application.[20]

The disagreement between Kuhn and Popper is partly over the question of whether or not a correspondence theory of truth is possible, Kuhn argues that it is not possible, and so we must rest content with a coherence theory: truth is an intra-theoretical term. But if this is the case, then any theory, domain, or field of human endeavor can be judged rational and true provided it has a paradigm, a method of completing the paradigm, and it is internally consistent. Rationality and truth, then, become ideals the satisfaction of which rests upon the prior acceptance of those assumptions that form the foundations of a domain. And when the strong fallibilist demands rational support for the assumptions themselves, champions of the coherence theory of truth are helpless. They have to resort to the kind of *ex cathedra* pronouncements that Kuhn provides, namely, that if science is not rational, then nothing is. The strong fallibilist rejoinder of course is to endorse the consequent of that dubious conditional.

Can Popper avoid this conclusion? If problems, the methods of solving them, and criticism are merely ways of getting from prescientific assumptions that form the traditional background to paradigms, then the ideal of objective truth recedes. The paradigm, it is true, may be approximated, but what is the guarantee that it is truth thereby which is being brought closer? How can it be determined whether or not the paradigm corresponds to reality? How could Popper avoid the possibility that a paradigm is an arbitrary way of looking at nature whose adequacy is judged by its internal consistency and the capacity to solve problems that arise only because scientists subscribe to prescientific assumptions?

A possible, albeit unsatisfactory, answer is provided by Popper's follower Lakatos, who frames his discussion in terms of research programs. Lakatos writes,

> One must never allow a research programme to become a *Weltanschauung* . . . setting itself up as an arbiter between explanation and non-explanation . . . . Unfortunately this is the position that Kuhn tends to advocate: indeed, what he calls "normal science" is nothing but a research programme that has achieved monopoly. But, as a matter of fact, research programmes have achieved monopoly only rarely and then only for relatively short periods . . . *The history of science has been and should be a history of competing research programmes (or, if you wish, "paradigms"), but it has not been and must not become a succession of periods of normal science; the sooner competition starts, the better for progress.*[21]

The guarantee that a paradigm does approximate nature, according to Lakatos, is that there are competing paradigms. The success of a research program means that it is a better model, a more accurate representation of nature than its competitor is. But how is the success of a paradigm or research program to be judged? One of Lakatos' examples of a research program is the mechanistic view of nature. Suppose it is contrasted with a teleological research program. According to one paradigm, nature is like a clock; according to the other paradigm, nature is like a growing acorn. Clearly, testing or prediction will not show what nature is like. If a scientist cannot find the causal law explaining the occurrence of an event, the mechanistic view of nature will not be abandoned. It will be assumed that the scientist has not found the law. Similarly, if the discovery of a fact seems to contradict the design that was thought to exist in nature, the conclusion will not be that nature has no design, it will be assumed that the design is other than what it was thought to be.

Research programs can be judged *vis-a-vis* the problems to which they attempted to provide a solution. Success consists in problem-solving, failure in the lack of it. As we have seen, however, problems that prompt research programs occur only against the background of expectations created by prescientific assumptions. Hence the adequacy of the research program must be judged against the rationality of those prescientific assumptions upon which it is based. And it is the rationality of these assumptions that the strong fallibilist calls into question. Lakatos recommends that scientists should build competing castles on air, and he ably reconstructs the architectural problems involved. The strong fallibilist is right in demanding something firmer

---

[20] T. S. Kuhn, "Reflections on my Critics," *loc. cit.*, pp. 265–66.

[21] See I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in CGK, p. 155.

V

If the arguments offered in this paper are correct, then there follows an old, but neglected problem: in order to maintain the rationality of science and in order to avoid strong fallibilism, the assumptions upon which scientific enquiry rest must be rationally supported. It does not seem to be questionable that there are at least some ontological, methodological, and categorial (classificatory) assumptions underlying scientific enquiry, and there may be axiological assumptions as well. It is only fitting that these assumptions be brought out into the open, examined, and given a chance to prove, as Popper would say, their mettle. The merit of fallibilism is calling attention to the problem. Popper fails in the attempt to deal with the problem by acknowledging that the assumptions are there, and insisting that they can, all the same, be subjected to rational appraisal by the use of the method of conjectures and refutations. For scientific enquiry, the practice of conjectures and refutations, owes its problems, methods, and ideals to the acceptance of those assumptions that it is now trying to make respectable. Unless independent, nonscientific, arguments are provided for the assumptions, science must be hopelessly question-begging. What could such independent arguments be? The answer is clear, although perhaps a little surprising: metaphysical. The conclusion seems to follow that the rationality of science and the possibility of metaphysics stand or fall together. The argument for this conclusion is the subject-matter of another paper.[22], [23]

University of Saskatchewan,
Regina Campus

Received October, 28 1971

---

[22] See "Metaphysics and Rationality," forthcoming in *Idealistic Studies*, *op. cit.*, 23.

[23] I am indebted to Donald F. Henze and Jean Y. Kekes for their comments on this paper.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 1972

IV. FICHTE'S IDEALISM

JOHN LACHS

FICHTE is usually classified as an idealist, yet the precise nature of his idealism is rarely examined. Idealism is frequently taken as the view that only minds and their states are real. The temper of our times is such that this theory appears to need no refutation. The suggestion that the Ohio Turnpike is nothing but a state of your mind or mine, or even of some supermind, no less than the Leibnizian idea that it consists of an infinite collection of harmonized souls, seem to us to be manifestly absurd and to warrant serious examination only of the persons who are so deluded as to suppose them true.

It would, of course, be wrong to classify Fichte as an idealist in any such popular or easy sense. He does maintain that there is a watershed question for all of philosophy: the answer to this determines the total complexion of one's philosophical thought. The question is whether or not objects exist independently of the self. There are only two possible answers to this question, and of the two only one is right. The dogmatist, incorrectly, asserts the absolute and independent existence of the world of objects. The idealist, by contrast, maintains that no object can exist independently of some self.

Fichte does not think that the world of objects, what we normally call the physical world, is in any sense the creature of some finite individual self, group of selves, or the infinite individual we call "God." Such might be the view of "dogmatic" or absolute idealism which, by Fichte's own profession, is in sharp contrast with what he calls his "critical idealism." According to the critical idealist both finite individual selves and the correlative world of physical objects are the resultants of a single, unconditioned self. Much of the interest of Fichte's metaphysics derives from his fundamental claim that the undifferentiated primordial being that is the source of all is actually a self. The plausibility of his idealism hinges entirely on his ability to substantiate this claim.

The central concept in Fichte's thought, then, is that of the self. What does he conceive to be the nature of selfhood? Without a reasonable account of this, his idealism is certain to remain vacuous. In attempting to determine his position on this and a variety of other issues I shall be relying primarily on the 1794 version of the *Wissenschaftslehre*, the first complete English translation of which was published last year.[1]

The most convenient and most obvious starting point in developing Fichte's thoughts on the self may well be his famous aphorism, "Was für sich nicht ist ist kein ich"—what is not conscious of itself is no self.[2] Self-consciousness is a necessary condition of selfhood: nothing that lacks the ability to reflect on itself can qualify as a self. Fichte affirms in a variety of places and in a variety of ways his conviction that self-consciousness presupposes consciousness of objects. This view may well be open to argument, but it is not my present purpose to dispute it. I shall merely note that a self must be or at least must have the ability to be conscious of both objects and itself. But even this is not enough for full-fledged selfhood: the intellectual, attentuated property of consciousness can never be enough to constitute the essence of a real being.

What, then, are the conditions both necessary and sufficient for a given existent's being a self? Are they to be consciousness and creativity? Or cognition along with a certain organization or structure of the elements of the being? Or the presentational and reflective activity of consciousness in conjunction with the purposiveness of all the activities of the being and the purposiveness of their union? Characteristically, Fichte has given ample thought to the problem, and equally characteristically he infuriates his readers by not disclosing his answer in any systematic or straightforward manner.

The magnitude of the reader's frustration may be evident from the fact that perhaps the best place for relevant hints on the nature of the self is in the chronically obscure Part III ("Foundation of the Knowledge of the Practical") of the *Wissenschaftslehre* of 1794. There Fichte discusses the activities of the absolute self in their purity; the activities and features that pertain to this primordial ego must be reflected, to a greater or lesser degree, in all finite individuals. The absolute self, Fichte maintains, has two divergent but interdependent fundamental drives. The two drives give rise to two divergent lines of activity; and since it is impossible to distinguish agent from activity on this level, the self is properly said to be constituted of these two sets of non-temporal acts.

---

[1] Johann G. Fichte, *The Science of Knowledge*, ed. and tr. by Peter Heath and John Lachs (New York, 1970). *Science* hereafter.

[2] *Science, ibid.*, p. 98.

The first activity of the self is its striving to fill out infinity. The practical drive urges the self on to infinite self-expansion, to an affirmation of itself and its own law without limit. The second activity is that of reflection or self-consciousness. The theoretical drive urges the self on to know itself as a unitary, self-expansive being. Using Fichte's physical model we could conceive the practical activity of the self as a line stretching outward from a center to infinity. By contrast, theoretical activity may be conceived as a line reverting back to the center; but throughout it is important to remember that the center is not to be thought of as any substantial or substantive being. The two activities are interdependent: reflection presupposes the outgoing activity which it restricts and turns back upon the self, while the effusive activity of self-assertion relies on the other to define its direction and to present it with the obstacles it must overcome.

These two closely interrelated activities define the nature of the self. Anything that is both self-assertive and reflective must be a self; nothing that lacks these characters can qualify as one. Reflectiveness is nothing but what I have previously referred to as self-consciousness and identified as a necessary condition of selfhood. If we combine it with expansive or assertive activity, we have the necessary and sufficient conditions of selfhood as Fichte conceives it. And this, incidentally, also gives us the clue to why Fichte thinks that, at least primordially, self-consciousness and the consciousness of objects are necessarily connected. The explanation is simple if we bear in mind that he thinks of reflection as an activity that restricts the infinite outgoing act of the self and drives it back upon itself. Since both activities are primordial, neither can overwhelm the other: no sooner is the assertive activity restricted than it renews itself, and the ensuing oscillation (Schweben), in a manner of speaking outside the self, yet totally dependent on it, is what Fichte calls the imagination. It is in this field of imagination that the presentations of which the empirical world consists are generated. The unavoidable conflict of the primordial activities manifests itself to the self in the form of the feeling of frustration and inability. These, in turn, are the source of the feeling of necessity or constraint which accompanies, in our experience, the presentation of external objects. In this way, the self's consciousness of itself is inseparable from the consciousness of external objects: the very activity of reflection is the cause, or at least a part of the cause, of the generation of objects.

What we have laboriously disentangled from the obscure parts of Fichte's work can also be found stated by him with the greatest clarity. Unfortunately, however, even the clearest and most elegant Latin is lost on the man who never learned the language. The meaning of some of Fichte's plainest statements becomes evident only after one has gone through some of his obscurest deductions: their clarity after one has understood them is often no less annoying than their vagueness had been before the light dawned. In a variety of places Fichte reminds us that the self is what it does, that it consists of nothing beyond its activities in their dialectical relation. If we add to this his oft-repeated statement that the self posits itself and posits itself as positing, we have in an embryonic form everything that we have said so far. For positing is nothing but the self-assertive activity of the self; and to posit the self as positing is to think it, to reflect on it, to know it as engaged in characteristic activity. The very essence of the self, therefore, is summed up by saying that it is something that both posits and knows that it does.

But have we helped clarify by introducing the concept of positing? Not if we can give no better account of the mysterious activity it stands for than the few commentators who have written on Fichte in English. Possibly the problem is one of translation: no English word captures the richness of the German original. The German word "setzen" is ordinarily translated as "to set," "to place," or "to establish." Its root significance is creative activity, an activity that can show itself in various modalities. It may be the simple physical act of placing an object in some location, the biological activity of bringing children into the world (Kinder in die welt zu setzen), or the exceptionally complex socio-political action of raising some person to the throne (auf den Thron setzen). What we have in each case is practical activity that is productive or creative; it is always purposive and often voluntary.

The activity "setzen" denotes, however, is not only practical. The word may be used to express agreement and opposition (sich auf etwas, wider etwas setzen), as well as the propositional attitudes of supposing and affirming. Its connection with the intellectual and the intelligible is further confirmed by the overtone we find of it in the German word for law, "Gesetz." The richness of "setzen" is due precisely to this amphibious character: it is equally at home in the realms of theory and practice. Fichte takes advantage of the ambiguity: he uses the word to denote an activity that is both cognitive and creative and represents the unity of reason and will, the theoretical and the practical.

The word "positing" is, therefore, at once revealed as inadequate to expressing what is conveyed by the word "setzen." "Affirmation" and "assertion" come close to capturing the volitional element in the activity Fichte refers to, and "self-affirmation" is certainly a central part of what he means when he says that the self "setzt" itself. But both the creativity and the cognitive action implied or suggested by "setzen" are lost in the word "affirmation," and we can safely say that there is no word in ordinary or, for that matter, in extraordinary English that comes any closer. What Fichte designates by this difficult word, then, is a fundamental cognitive-conative activity. It is a purposive and productive act, an act whose creativity is the source of all that is real. If we keep this in mind, it comes as no surprise that Fichte thinks the organ of positing is reason itself. Reason in its primordial unity is thus conceived as the infinite and intelligent source of all, totally absorbed in its creative, all-encompassing act. For lack of a better substitute, I shall continue to use the word "posit" to stand for Fichte's "setzen." But it is important to remember that I shall refer by this word to the cognitive-conative activity whose nature I have briefly indicated.

Positing is a non-temporal act. Its model no doubt is the Aristotelian concept of activity or "energeia" in which process and product, agent, act, and deed indistinguishably coincide. This concept of activity appears in Spinoza, whom Fichte frequently praises as the greatest and most consistent of dogmatists, as the infinite potency-in-act that is his all-creative Substance. Viewed dynamically, from the standpoint of the act, Substance or God is *Natura naturans*, an indeterminately infinite creative agency. Viewed statically, from the point of view of the deed, Substance or God is *Natura naturata*, an infinitely determinate, eternal modal order. Since productive act and completed deed are indistinguishable in the primordial, infinite potency-in-act and since no temporal lag separates activity from its end, creative energy and created world are inseparably one.[3]

I do not wish to minimize the difference between Fichte and Spinoza. But I also do not wish to overlook or underplay instructive similarities. Positing, as the absolute ego's primordial activity, is clearly analogous to what in Spinoza I have called potency-in-act. And the relation of Spinoza's God as creative act to God as created modal order has a clear analogue in the relation, in Fichte, of the absolute ego to the empirical world of subjects and objects. The progression from indeterminate infinity to infinite determinateness is present in both systems: it is by this gradual determination that, in both, finite fragmented beings are generated. And there is agreement that the progressive self-determination of the absolute is a necessary consequence of its nature. The two thinkers are in consonance even on the fundamental neo-Platonic dictum that determination is negation: the self-determination of the absolute, by its application to itself of ever more specific predicates, is at once a self-negation and self-limitation.

We may view Fichte's interesting theory of thetic judgments in the light of its relation to the emanation scheme that is one of his neglected, or at least little understood, versions of the generation of everything finite from the infinite ego. A thetic judgment, according to him, is one that consists of the affirmation of the existence of a subject without any reference to a predicate. The prime example he gives is the judgment "I am." Now it is clear that the being of the absolute self would have to be affirmed in a thetic judgment, and the peculiarity of such a judgment is that it leaves the concept of its subject unrelated to any other concept, and thus totally indeterminate. This is the infinite indeterminacy of the absolute ego; yet on sooner is the indeterminacy ascribed to the self than it becomes inapplicable. Even as general a predicate as indeterminateness limits and determines; the very indeterminateness of the primordial self constitutes the first step toward its inevitable total determination. The finite world of empirical selves and physical objects naturally flows from the necessary self-determination of the absolute ego, and in Part II ("Foundation of Theoretical Knowledge") of the *Wissenschaftslehre* of 1794 Fichte displays considerable ingenuity in tracing this dialectical development.

---

[3] I acknowledge my indebtedness for this interpretation of Spinoza to H. F. Hallett's excellent *Benedict de Spinoza* (London, 1957).

An obvious dissimilarity between Fichte and Spinoza, of course, is due to the historical fact of the presence of Kant between them. Spinoza strove to deduce the eternal modal order from the activity of his primordial substance. By contrast, Fichte no longer wishes to deduce an infinity of beings, but only the structures and categories of experience. I will not stop to explore this interesting point; it is peripheral to my main purpose here. I have set out to examine the nature and to determine the tenability of Fichte's critical idealism, and I have now managed to reach a stage where his terms have been, I hope, adequately clarified, and where we can set him a fundamental and, I think, disastrously destructive question.

Fichte spends considerable effort in deducing the categories, conceived as principles of unity in experience, from the self-postulation of the absolute self. It should be evident, therefore, that the primordial self itself is sub-categoreal or pre-categoreal. We cannot say of it, for example, that it is a substance, even though we must think of it as, in a sense, non-temporally enduring. No predicate that is applicable in the world of experience is applicable to it univocally; in its primordial being, which may be affirmed in a thetic judgment, no predicate is applicable to it at all. I shall not discuss here Fichte's apparent violation of this principle in conceiving of the absolute self as the cause of the world of objects. Spinoza at least realized a part of the limitation on what we may predicate of the primordial being and conceived his infinite potency-in-act as neither mental nor physical, although in a sense the source of both. The question fundamental for Fichte's idealism that thus arises is the following: What conceivable reason can we offer for thinking of the primordial agency-activity as a self? If there is good reason for thinking of it that way, it may be difficult to escape Fichte's critical idealism; if there is none, his central views will surely appear as unwarranted dogmas.

To say, as Fichte does repeatedly, that the absolute ego is not in any sense an individual person does not eliminate the question. On the contrary, it only raises the additional problem of how anything could be a self without displaying at least some of the determinate tendencies, attitudes, and intentions that are the hallmarks of individual personality.

I think I can detect in Fichte three major lines of argument for his idealistic conclusion. Nowhere does he develop these arguments in detail. Their outlines, however, are clear enough and he hints at them in a variety of places. I shall consider them in turn.

The first argument starts from the premiss that the task of philosophy is to give an account of the origin and nature of experience. Fichte appears to be of the opinion that this can be attempted in only two ways. In the first attempt we proceed from the side of the object or unthinking thing, in the second from the side of the subject or self. Accordingly, as he insists in the First Introduction to the *Wissenschaftslehre*, there are only two possible philosophical positions: dogmatism, which attempts to deduce experience from a world of independently existing things or things-in-themselves, and idealism, which attempts to display experience as the result of the operation of an active self.[4] The two theories are contradictory: at most one of them can be true and at least one of them must be. Fichte makes no secret of the fact that he thinks we can never account for lived experience by reference to things or objects alone. In some places he goes so far as to assert that a dogmatism which attempts such an account is demonstrably false. If only two attempts at explanation are possible, the two are contradictory, and one of them is demonstrably inadequate, the other must clearly be correct. For this reason, idealism is the only tenable philosophy and the primordial existent must be a self.

This argument appears to me to have no merit. First of all, even if we grant that one of the tasks of philosophy is the relatively vague one of explaining the nature and origin of experience, we have reason to doubt the remarkable claim that this can be done in two ways only: either by sole reference to things and their laws or by reducing all to selves and their operations. This is surely a gross simplification and excludes by edict far more philosophical views than it permits. It is this oversimplified picture that leads Fichte to the mistaken thought that idealism and dogmatism are true contradictories and that hence by disproving the latter we can incontrovertibly establish the former. The contradictory of the proposition that in explaining the source and structure of experience we need no reference to anything beyond things and the laws that govern their behavior is the proposition that such reference is necessary, and not the proposition that we need reference to selves only. Even if we were to stay within Fichte's unreasonable thing—self dichotomy, the most that the refutation of dogmatism could show is that some concepts or laws other than those relating exclusively to things are necessary to give an adequate account of human experience. This would leave the door wide open to a variety of dualisms, among them the highly critical and attenuated sort held by Kant, who viewed experience as a unity of elements derived from self and thing.[5]

---

[4] *Science, op. cit.*, pp. 9ff.

Let me wipe the slate clean, however, and permit the entire matter to be decided by an adequate refutation of dogmatism. On Fichte's own interpretation such a demonstration of the inadequacy of dogmatism would be tantamount to a proof of his idealism and even on a less charitable rendering it would tend to increase substantially the probable truth of his position. Unfortunately, however, one looks in vain through Fichte's works for the demonstration. It is, in fact, not even seriously attempted, much less accomplished. And in a bold about-face in the work in which he claims the demonstrable falsity of dogmatism, Fichte asserts that ultimate philosophical positions are susceptible neither of proof nor of disproof: the view we adopt, without evidence and as a starting-point, is a function of personal interest, commitment, and disposition.[6] I shall come back to this point later and now proceed to Fichte's second argument.

Here his reasoning amounts to perhaps no more than a version of the previous argument stated in the material mode. Whatever else there may be in the world, it is generally agreed that there are at least some selves. Now Fichte thinks it evident that no self can ever come from an unfeeling, unthinking not-self. If all is to have a simple, unitary source and the only conceivable source of an ego is another one, it clearly follows that the primordial reality must be a self. We can view Fichte's attempt to deduce finite objects and subjects from an absolute self as giving added point and poignancy to this argument. Since nothing that lacks self-consciousness and self-affirmation can be the ground of a self, nothing that is bereft of these features of selfhood can serve as an ultimate principle. His deduction aims at establishing that an ego, by contrast, is able to generate both selves and things, and thus amply qualifies as the ultimate source of all.

I cannot make myself believe that this argument has anything to recommend it. First and foremost, I am entirely at a loss as to why anyone should think it evident that a self cannot have its source in the not-self. I do not propose to argue here that it can or does, even though the entire weight of evolutionary thought supports this hypothesis. I shall content myself with the comment that if no good reasons are presented for accepting it, Fichte's putatively self-evident premiss is simply an arbitrary dogma. The issue of whether or not things can function as the source of selves is at the very heart of the problem of idealism: the bold claim that they cannot serves only to start the controversy, not to resolve it.

Fichte was no less vocal in avowing belief in the view that things cannot give rise to selves than he was silent about his reasons. Although it is idle to speculate on what these might have been, it is appropriate to make two brief remarks. The first is that his reasons could not have been of the sort that supported Berkeley's conviction that objects cannot generate spirits. For Berkeley thought of causation as the exclusive property of minds; for him physical objects were impotent. By contrast, Fichte in no way restricts causal activity to selves: following Kant he thinks of causation and of reciprocity as universally applicable. My second and more positive comment is that what might be behind Fichte's conviction that the not-self can generate no self is simply the time-honored causal maxim that decrees the impossibility of the like coming from the unlike. The putative emergence of mind, consciousness, or selfhood from the thoughtless and inanimate is surely a paradigm of the generation of the like from the unlike. To some it may also seem to demonstrate the absurdity of any such generation. In reality, however, there is nothing absurd about it, and if Fichte's argument is to be grounded in this maxim of causation, we are perfectly justified in asking for the grounds for it. The blow of the hammer does not resemble the pain it causes, nor the brain of Shakespeare a page of manuscript. The vast majority of the causal sequences we know involve the creation of like by unlike; what are we to do with a maxim that decrees all this impossible?

---

[5] Fichte devotes considerable effort and ingenuity to the attempt to show that Kant held no such view.

[6] *First Introduction* to the *Wissenschaftslehre, Science, op. cit.*, pp. 12–16.

Let me, however, forget all that has been said: let us suppose that Fichte has proved, as he has not, that the source of every self is itself or another. Would this be adequate to show that the ultimate existent must be an ego? Clearly not. This by itself leaves the door wide open, once again, to dualism: it is perfectly compatible with the view that there are two ultimate existents, say, a transcendental ego and a transcendental object. Fichte thus needs an additional principle, viz., the one that maintains that all is to have a single source. Yet why should we accept this principle? Surely not because unity is boldly asserted to be a demand of reason. This may well be the demand of reason in a tight little mind perpetually at work imposing discipline on his children or the world, but it is surely no demand of universal reason. And even if it were, would it not be a begging of the question, an unwarranted assumption of the idealist principle that reason shapes the world—the very principle under examination—if we were to say that its demands are invariably met?

I now come to Fichte's third argument. This is never explicitly stated by him, but it appears to me to be behind what he says in a number of passages. We start with the assertion that the primordial being must engage in creative activity. Such an activity, since it is essentially self-assertive and expansive, resembles far more closely the activity of a self than that of some thing or not-self; self-assertiveness is in fact, as we noted, one of the marks of the ego. If the argument were correct, it would not of course prove that the ultimate source of all is actually a self. It would, however, show that such an unconditioned being must resemble selves far more closely than it resembles unconscious or inactive things—and this may conceivably be all that Fichte needs to establish his idealism.

This argument hinges on a fundamental unjustified assumption, namely that the activities of the self are the paradigm of creative activity. Fichte clearly wants to stop short of the extreme position of asserting that only selves can engage in causally creative endeavor; since he does, he has to justify his taking the dialectical self-expansive/self-limiting activity of the self as the model of all creative agency. Why should we say that the primordial creative act is more like the self-expansiveness of a self than like the self-expansiveness of a gas or of a healthy tree? We search in vain for Fichte's answer to this question. It is clear that reference to the self's voluntary or intelligent self-assertiveness is of no avail: the absolute ego's expansiveness is unavoidable and the self-expansion of a tree is law-governed and purposive. Similarly, there is little merit in the approach that professes to see in each tree's growth a pale replica of the ego's actions. These activities are different, and either one can be considered a pale or imperfect version of the other, depending on what features we select for emphasis. Fichte loads the dice by using the word "setzen" for the primordial activity: this suggests, without being evidence for, the view that the activity properly belongs in the sphere of the self. But what reason have we to suppose that the activity is self-like or self-connected and that "setzen" is not merely a misleading or question-begging name for it? As best I can see, none at all.

I must conclude, therefore, that Fichte's arguments for the view that the primordial being is a self fall altogether short of establishing their conclusion. As we have seen, each suffers from a variety of faults, but most blatantly the first fails because it presupposes an unprovided demonstration, the second because it must rely on an unacceptable maxim of causation, and the third because it begs the question by a prejudiced selection of the paradigm of activity. Does this leave Fichte entirely without defense? I do not think so. It would be open to him to use a version of his third argument, if not to attempt to show that the ultimate being is a self, then at least to justify supposing it to be that for the sake of increasing our knowledge. This regulative version of the argument is well known in the history of idealism, and the likely reason why Fichte made no reference to it is that he thought it both valid beyond all reasonable doubt and inextricably intertwined with the entire enterprise of Kantian and post-Kantian transcendental philosophy. The usual argument is that since we know the self and its structures and activities best of all, we must use them as our model for understanding everything else. Fichte, however, might subscribe to a somewhat stronger version. His acceptance of Kant's theory that all intelligible structure is due to mental activity might lead him to the view that all we ever know is the self, its structures and activities and therefore have in fact no choice but to use them as our models.

Now I can find nothing wrong with using the self as our model in the attempt to see how far it will take us and what important or unsuspected insights it might provide. One can have one's opinions about the ultimate value of this model, but when all is said and done there is no substitute for the actual full-scale attempt to push the idealist program as far as it will go. Surely this is all the justification Fichte needs for having attempted a deduction of the finite world of subjects and objects from a primordial self. It is an interesting and important fact that such a deduction is possible, and his having managed to complete it is a tribute to his dialectical skill. But the possibility of the deduction is one thing, its truth or bearing on the real world another. It is well worth keeping in mind that the way we choose between two deductions, say those of Spinoza and Fichte, is normally not by finding logical or dialectical slips in the one and none or fewer in the other. Philosophers notoriously have more logical than good sense, and though an idealist deduction proceeding on the basis of the regulative principle under discussion is clearly possible, it may well be severely at odds with reality.

So much for the weaker version of the regulative principle. The stronger version, according to which our only models for understanding the world are the cognitive, conative, and affective activities of the self appears, by contrast, excessive and question-begging. It assumes, without good reason, that all forms, structures, and activities are mental. By according such pre-eminence to the self and thinking of it as the source of all order, if not of all reality, were we not tacitly affirming idealism? And if such tacit affirmation is disavowed, what reason could conceivably be proposed for saying that the activity of gravitation or the orderliness in the growth of poison ivy reveal somehow the marks of intellect?

Let me conclude with some remarks on a characteristic, though typically exaggerated, claim of Fichte's. Acceptance of ultimate philosophical positions is a matter of personal inclination or interest, he maintains. Now, his argument continues, the highest and most exalted interest of every ego is that of self-affirmation or the free exercise of the intelligent creative powers of the self. This activity of self-creation and self-development is the source of all morality and goodness. But an activity of this sort presupposes a free and independent ego. The highest interest of man, the very possibility of morality demands, therefore, the primacy of the ego and its total primordial independence of determination by objects. This makes idealism, and with it belief in the primacy of the self, a demand of morality, Fichte concludes. But he refuses to stop even here and indicates that whoever does not acknowledge and exercise his freedom does not truly or for long possess it. And not to possess freedom is not to be a self, but a mere thing. For this reason, failure to subscribe to idealism is not merely a theoretical error: it is a practical act which disqualifies one from membership in the kingdom of ends. The dogmatist is not a person who is simply wrong; he is simply not a person.

It is difficult to think that this argument is anything but nonsense. It is a peculiar sort of morality indeed that sees the greatest requirement imposed on every man as self-assertion or the exercise of his native powers. But this peculiarity is nothing when compared to the singular claim that practical life determines or should determine one's beliefs about what is, and that a single wrong opinion can strip us of selfhood and turn us into things.

The ultimate problem here, however, is the easy confusion between the absolute ego and the finite selves of individuals. Fichte devotes considerable attention to the deduction of the world of objects, but almost none at all to the deduction of finite subjects. To establish his idealism, he would have to prove that the unconditioned being is a self. Showing, as he attempts to, that the demand of morality is to think of individual selves as unconditioned and free does not in the least contribute to proving this. Good sense impels us to believe in the existence of a variety of finite selves; but if there are more than one of them, none can serve as the primordial unity from which all flows. If, however, all individual selves are to share a single transcendental ego, the analogy, central to Fichte's entire enterprise, between our finite selves and this primordial ego inevitably collapses. And even if this near disaster could be averted, what account should we give of the relation, so peculiar on this view, of transcendental to individual self? To this question, though not to this alone, Fichte provides no answer.

Vanderbilt University

Received October 28, 1971

BIBLIOGRAPHY

Fichte's works are readily available in German. His son, I. H. Fichte, brought out an edition of his collected works in 1845–1846 (*Sämmtliche Werke.* 8 vols. Berlin, Veit). This has now been supplanted by the *variorum* Gesamtausgabe of the Bayerischen Akademie der Wissenschaften (*Werke.* Stuttgart-Bad Cannstatt: F. Fromann, 1964–). There is a readily available reprint of *Grundlage der Gesamten Wissenschaftslehre* (1794) by F. Meiner Verlag, 1961. No English edition of the collected works exists. A substantial number of Fichte's epistemological and metaphysical writings remain, in fact, untranslated. An incomplete and inaccurate rendering of the *Grundlage* appeared in 1889 (*The Science of Knowledge.* Translated by A. E. Kroeger. London, Trübner). A new, complete translation (J. G. Fichte, *Science of Knowledge*, edited and translated by Peter Heath and John Lachs. New York: Appleton-Century-Crofts, 1970) incorporates also the First and Second Introductions to the *Wissenschaftslehre.*

There is a substantial body of scholarly and critical literature on Fichte in German. The following list is a selection from those most interesting or most relevant to the thesis of this article.

Friedrich Dannenberg, *Der Begriff und die Bedeutung der Erfahrung in der Fichteschen Philosophie.* Weida i. Th.: Thomas and Hubert, 1910.

Julius Drechsler, *Fichtes Lehre vom Bild.* Stuttgart: Kohlhammer, 1955.

Kuno Fischer, *Geschichte der neuern Philosophie.* Heidelberg: C. Winter, 1897–1904. Vol. 6.

Dieter Henrich, *Fichtes ursprüngliche Einsicht.* Frankfurt a.M.: Klostermann, 1967.

Willy Kabitz, *Studien zur Entwicklungsgeschichte der Fichteschen Wissenschaftslehre aus der Kantischen Philosophie.* Darmstadt: Wissenschaftliche Buchgesellschaft, 1968.

Dietrich H. Kerler, *Die Philosophie des Absoluten in der Fichteschen Wissenschaftslehre.* Ansbach: C. Brügel, 1917.

Richard Kroner, *Von Kant bis Hegel.* Tübingen: Mohr, 1961.

Johann H. Loewe, *Die Philosophie Fichtes: Nach dem Gesammtergebnisse ihrer Entwicke lung und in ihrem Verhältnisse zur Kant und Spinoza.* Stuttgart: W. Nitzschke. 1862.

Fritz Medicus, *J. G. Fichte: Dreizehn Vorlesungen.* Berlin: Reuther & Reichard, 1905.

Baldwin, Noll, *Kants und Fichtes Frage nach dem Ding.* Frankfurt a.M.: Klostermann, 1936.

In English there is very little critical literature on Fichte's metaphysics that is worth reading. The following four books may be consulted, though without any great hope of enlightenment:

Robert Adamson, *Fichte.* Edinburgh and London: W. Blackwood, 1881.

Charles C. Everett, *Fichte's Science of Knowledge: A Critical Study.* Chicago: Griggs Philosophical Classics, 1892.

Ellen B. Talbot, *The Fundamental Principle of Fichte's Philosophy.* New York: Macmillan, 1906.

Anna B. Thompson, *The Unity of Fichte's Doctrine of Knowledge.* Boston: Grinn & Co., 1895.

Among histories of philosophy in English, by far the most adequate account of Fichte is in F. H. in Coppleston, *A History of Philosophy* (Garden City: Doubleday, 1965. Vol. 7, Part I). The article in P. Edwards' *Encyclopedia of Philosophy* is virtually useless. The most recent and most interesting English treatment of the *Wissenschaftslehre* of 1794 is an unpublished doctoral dissertation (Walter E. Wright, *Self and Absolute in the Philosophy of Fichte*, Vanderbilt University, 1971).

I shall mention only two recent works in French:

Xavier Léon, *Fichte et son temps.* 2 vols. Paris: Armand Colin, 1954.

A. Philonenko, *La Liberté Humaine dans la Philosophie de Fichte.* Paris: Vrin, 1966. (A very thoughtful work.)

On the whole, however, one who is interested in Fichte's systematic philosophy will do best by proceeding to the texts directly. A remarkable scholar of the stature of Kuno Fischer can help to put Fichte in the proper historical perspective. But, alas, in dealing with the text one seeks in vain the aid of even one commentary as careful and sustained as exist in droves on the work of other major thinkers.

American Philosophical Quarterly
Volume 9, Number 4, October 1972

V. NON-CONSCIOUS SEEING

PATRICK McKEE

I

One kind of non-conscious seeing occurs when a motorist, driving on an unfamiliar road and unaware of a red traffic light in his field of view, drives past that red traffic light, immediately afterwards comes to believe that he has just driven past a red traffic light, and is then able to say, in defending his belief that he has just passed a red light, that he can remember that the thing *looked* like a red light, including its having looked red to him. Call the motorist *M*, the time at which the light was in his field of view t_1, and the time, subsequently to t_1, at which he comes to believe that he has just passed a red light, and is able to say, on the basis of memory, that the light looked red to him, t_2. We must concede that *M* was not aware of the light at t_1 because he explicitly remembers that his thoughts were entirely taken up with other things at the time, and explicitly remembers not being aware of the light. Nevertheless, because of *M*'s remembering that the light looked red to him at t_1, we must concede that it looked red to him at t_1.

Among the questions that now arise are: How is it possible for one to be in a position to say "It looked red" in such cases? Don't we need to explain how it is possible for *M* to say, on the basis of memory, that at a certain moment in the past when he was not aware of a thing that it was, nevertheless, then looking some way to him?

One answer that suggests itself is that something happened in *M*'s mind at t_1, namely, a certain subconscious, inner mental event—a subconscious visual sensation—occurred, and that *M*'s bringing that mental event to consciousness in memory is what makes it possible for him to say, at t_2, "It looked red." But there is an argument against this suggestion. Suppose it were correct that what makes it possible for *M* to say "It looked red" is his remembering an "inner occurrence" of a visual sensation. Then it ought to be possible for him to try to remember the visual sensation quite precisely. One can always be said to try to remember precisely, and then quite precisely, anything that one can remember having experienced. Now consider the target expression "It looked red." If one attempts to remember *quite precisely* what the "visual sensation" was like in such cases of seeing, one's memory focuses instead on the conscious thoughts and feelings one had at the time or on what was happening at the time, or on the movements one made, and so forth. And notice that the desire to "remember more precisely" what the seeing was like in such cases is *satisfied* by remembering these things! So if we construe the remark "It looked red" as involving reference to an inner sensory event, the inner sensory event apparently drops out of consideration as irrelevant.

Though the light's looking red to *M* did not include *M*'s awareness of the light's looking red, nevertheless what I am calling "the light's looking red to *M*" did not consist, solely, in *M*'s disposition to say "Yes," if asked, at the time of his seeing, "Does that light look red to you?" or in *M*'s disposition to say "It looked red" after the time of his seeing, or any other of *M*'s dispositions. For *M*'s belief, adopted at t_2, that the light was red at t_1, is a justified belief, whose justification is directly related to the light's looking as it did to him at t_1.

That *M*'s seeing did not include awareness of the light's looking red to him, but did involve more than *M*'s related dispositions, suggests the presence of some intermediate state—something between awareness of appearances on the one hand and disposition on the other. To see the need for the concept of such an intermediate state, consider the account of non-conscious seeing given by Fred I. Dretske in his book *Seeing and Knowing* (Chicago: University of Chicago Press, 1969). According to Dretske the seeing featured in such cases is "non-epistemic" seeing. The percipient sees, but in seeing acquires no beliefs, and specifically does not acquire the belief that something is appearing in some way to him (pp. 11–12). Nevertheless, this seeing includes the object's looking some way to the percipient (p. 22). Now the trouble is, Dretske cannot characterize the looking included in non-conscious seeing as including the percipient's awareness that he is being appeared to in some way, for then knowledge of, and some belief about, his being appeared to is entailed, and it is just this entailment Dretske argues against, for such cases (pp. 11–12). Yet if one denies that the percipient is aware of being appeared to at the time of seeing, that suggests that an object's looking red to a percipient is a matter of mere discrimination, as a behaviorist might characterize "visual differentiation." But Dretske can't take this suggestion either, because for him the appearing included in non-epistemic seeing is the sensory element or dimension of perception, and serves as grounds for perceptual belief (pp. 75–76; pp. 79–81). But if the being-appeared-to featured in non-conscious seeing is not a form of sensory awareness, and yet is more than a physiological state apt for causing discriminatory behavior, what's in between these two? In what follows I attempt to bridge this gap by introducing the notion of "explanatory sensation." I shall proceed by first explaining briefly what I mean by explanatory sensation, then developing this concept in detail by tracing its connection with our ordinary way of thinking about non-conscious seeing.

II

According to one way of thinking about appear-verbs and other psychological verbs, they must always describe mental acts, processes, or states. This doctrine of the psychological is usually accompanied by a certain picture of what an event, process, or act is, such that a thing's having looked ϕ to someone at time t_1 must be temporally discrete, in the sense that it either looked ϕ to that person at t_1 or else it didn't, and its looking ϕ to him, if it did, and its not looking ϕ to him, if it didn't, is independent of what comes after t_1. Now this way of thinking does not seem to be true of every past-tense use of psychological verbs. If a man stumbles at the top of the stairs and is asked why, his reply "I thought there was another stair" does not (necessarily) describe a discrete event-thought which he had at the time either in the sense of forming the words "there is another stair" or in the sense that his mind was engaged in entertaining the proposition "there is another stair" before or at the time of his stumbling. If one's saying, upon stumbling, "I thought there was another stair" were a report of a discrete event-thought, then we ought to be able to imagine just that same thought by itself, cut out from antecedent and subsequent developments. But we can't imagine this; cut off from its antecedents and its development no occurrence or isolated episode would be that of his having *that* thought of there being another stair. Nothing would count as the thought of there being another stair which we are now concerned with apart from his climbing the stairs and stumbling. What would it be like to take away the antecedent climbing of stairs and the stumbling and leave a mental occurrence corresponding to his having that thought "There is another stair"?

In the characteristic use we are examining the expression "I thought there was another stair" is not a report of a prior discrete mental event, but an explanatory remark. I make the remark "I thought there was another stair" to explain why I stumbled, to make it clear to myself and to others why I stumbled. Now it follows from what has already been said that the explanation this remark gives does not depend upon its adverting to a prior discrete event-thought. But that does not mean that there was no thought, that really I *didn't* think there was another stair. Our ordinary way of thinking about such cases embodies an alternative conception of this "thought"; namely, that my subsequently being able to explain my stumbling (in a sense of "explain" to be explicated in Sect. III, below) by saying "I thought there was another stair" is what it is for me to have had the thought then—when I stumbled—that there was another stair. Against this we may feel entitled to insist that if a man had a thought at a certain time then either he *had* that thought, or else he *didn't*, and his having had the thought, if he did have it, and his not having it, if he didn't, is independent of subsequent developments. We seem to want to insist that statements asserting the existence of past thoughts must be categorical, in the sense that their truth cannot be conditional upon subsequent developments. I want to suggest that this insistence is alien to our ordinary way of thinking about such cases. The ordinary way of thinking about such cases does not endorse either a categorical "Yes" or a categorical "No" in answer to the question whether there really was a thought. In answering the question "Did the man who stumbled on the stairs really have the thought that there was another stair?" we find it necessary to make reference to the situation in which his stumbling occurred and to our way of explaining and understanding that stumbling. To the extent that we find the remark "He thought there was another stair" necessary for understanding his stumbling, the answer to the question whether there was a prior thought is affirmative. And to whatever extent (if any) reference to a prior thought might be shown to be not necessary to our understanding his stumbling, our ordinary way of thinking about the case condones our relinquishing the existence-claim that there was such a thought. That is to say, our being able to best explain the stumbling by saying "He thought there was another stair" is what it is, according to our ordinary view of such cases, for there to have been that thought. We are committed to the existence of that thought precisely and only to the extent that we accept that advance in our understanding of the situation that comes with the man's remarking "I thought there was another stair" and by others' remarking "He thought there was another stair." Accordingly, I shall refer to such thoughts as "explanatory thoughts."

Now consider again the traffic light's having looked red to *M*. I shall adopt the hypothesis, subject to withdrawal upon its failure to be confirmed by what follows that just as one's saying, in explanation of his stumbling, "I thought there was another stair," does not entail that one's mind was engaged in entertaining the thought "There is another stair," but rather his being able to explain his stumbling in this way is what it is for him to have had the thought, so it is with the light's having looked red to *M*. That is to say, *M*'s saying, in connection with the kind of seeing which now concerns us, that the thing he saw looked red to him, does not entail that in addition to his job of shifting gears and pedalling the accelerator, *M* was also engaged in appreciating the look of the light. Rather, the light's looking the way it did to *M* at t_1 consists in his being able, at t_2, to offer the statement "It looked red" in explanation of his believing that there was a red traffic light at the intersection. When a thing's having looked ϕ to someone is, simply, for him to be able to advert to that thing's having looked ϕ to him in explanation of a belief he has adopted, I shall refer to its having looked ϕ to him as an "explanatory visual sensation." (In this connection I use the term "sensation" as a label only; in the arguments that follow I do not appeal to any use of this term or any natural or philosophical conceptions corresponding to it.)

III

But if *M*'s statement "It looked red" explains his belief that there was a red traffic light at the intersection, what is the logic of the explanation it gives? It might be explicated, in accordance with an inner-event picture of visual sensation, as follows:

> In saying "It looked red" *M* explains his belief that there was a red traffic light at the intersection, in the sense of making it clear why he has come to believe there was a red traffic light at the intersection, by asserting the existence at t_1 of a remembered visual sensation consisting in the light's looking a certain way to him, whose existence at t_1 is logically independent of *M*'s adopting that belief at t_2, and is, according to a "rule of evidence," sufficient to render that belief rational for *M*.

As an analysis of the explanation of *M*'s belief provided by his saying "It looked red" this is at best only a beginning, since *M*'s remembering a thing's having looked red to him in the indeterminate way he now remembers the traffic light's having looked would not by itself be sufficient to confer rationality on his present belief that there was a red traffic light at the intersection. For even if we imagine *M*'s remembering not only a visual sensation as of an (indeterminately) red thing' but as of that thing's appearing (indeterminately) of a certain size and shape, and at a certain distance and direction from his eyes, still his now remembering this visual sensation would not by itself be sufficient to confer rationality on the belief that there was a red traffic light at the intersection. Beyond having had such a visual sensation, *M* must, presumably, on this account, have certain "background beliefs," such as that the conditions of observation for driving were normal at the time. If, for example, *M* believed that the lighting fixtures and sign-posts of the intersection were decorated by red Christmas street-lights, then a thing's having looked red and so forth in the indeterminate way he now remembers would not be sufficient to confer rationality on his belief that there was a red traffic light at the intersection. For parallel reasons *M* must, presumably, have the background beliefs that traffic lights are of a characteristic size, and are stationed at characteristic positions in intersections. Might *M*'s saying "It looked red" then explain his belief that there was a red traffic light at the intersection by adverting to a visual sensation, consisting in the light's looking as it did to him, and which is, in accordance with a rule of evidence, a sufficient condition for the belief's being rational when he believes these other things?

This way of thinking cannot adequately

explicate the explanation provided by *M*'s statement "It looked red." First, we accept *M*'s statement as providing the explanation it does when we do not know he has the relevant further beliefs, in the sense that the question of what beliefs accompany his contention that what he saw looked red simply doesn't arise. Secondly, and more fundamental, according to this explication the further beliefs in question are conceived, as is the light's looking red to *M*, as existing independently of the belief to be explained. But this is confused; the beliefs in question are (typically) specimens of "explanatory" thoughts. For consider: *in what sense* does *M* think, at the time he says "It looked red," that the conditions of observation for driving were normal, or that traffic lights are stationed at such and such positions in intersections? Not in the sense that he forms words to that effect, or entertains propositions to that effect, but rather in the sense that if his belief that there was a red traffic light at the intersection proves false (he discovers that what he saw was a decorative Christmas-light), he could nevertheless defend the reasonableness of his prior belief by saying "I'd forgotten about the decorations—I thought the lighting was normal" (or something of the sort), or if his belief that what he saw was a red traffic light were criticized on the grounds that traffic lights are never placed to the right of intersections but always suspended overhead in the middle of intersections, he could nevertheless defend the reasonableness of his belief that what he saw was a traffic light by saying "I thought traffic lights were sometimes placed over to the right." So "background beliefs" enter into the explanation, if at all, not as logically independent conditions of the rationality of the belief to be explained, but as logically dependent conditions for its rationality—i.e., only insofar as the belief's rationality has been called into question and can now be made clear by *M*'s saying "I forgot about the decoration lights," and the like. This leaves only the light's looking the way it did as candidate for independent sufficient condition for the reasonableness of the belief. But we have already seen that the light's looking as it did cannot alone be sufficient. What is needed, apparently, is an explication of the explanation provided by *M*'s saying "It looked red" which does not depend upon construing the relevant sensory event consisting in the light's having looked as it did to *M* as an independent sufficient condition for the rationality of his belief.

No doubt the explanation of *M*'s belief provided by his statement "It looked red" consists in making clear the rationality of his belief, enabling us to regard it as rational. This understanding of *M*'s belief as a rational belief, effected by his explanation, consists in our recognizing, upon hearing his statement "It looked red," that *M* has a reason for adopting the belief that there was a red traffic light at the intersection. His statement provides the understanding it does by specifying his reason for belief. But how can we make analytical sense of this understanding, given that it cannot involve recognition of a sensory event which is a sufficient condition for the rationality of *M*'s belief? The needed explication will not become clear until we separate the following as distinguishable aspects of the understanding *M*'s statement provides: (1) our recognition of what kind of belief *M*'s belief that there was a red traffic light at the intersection is; (2) that *M* has a reason for adopting it; and (3) what that reason is.

We recognize what kind of belief *M*'s belief is, namely, perceptual belief, in recognizing that *M* takes himself to have acquired it through the use of his senses, and specifically through the use of his eyes. Further, *M*'s description of how the thing he saw looked to him makes it clear that *M* has a reason for adopting the belief that there was a red traffic light at the intersection. But that aspect of our understanding of *M*'s belief consisting specifically in our recognizing that *M* has a reason for adopting the belief that there was a red traffic light at the intersection does not itself presuppose anything about what that reason is. It does not, in particular, presuppose that *M*'s reason for adopting the belief he does is the light's having looked in some way to him. This is shown by the possibility that another percipient (let us identify him as *N*), upon passing the red light at the intersection without having been aware of the light at any time while it was in his field of view, might adopt the same belief as *M*—the perceptual belief that there was a red traffic light at the intersection—on the basis of some reason, but without the light's having looked in some way to him. Like *M*, *N* believes that he has acquired the information that there was a red traffic light at the intersection, and that he has acquired that information through the use of his eyes. Discounting as differences in the content of the two men's beliefs any differences consisting merely in *M*'s beliefs referring to *M* and *N*'s referring to *N*, we can imagine their beliefs to be identical in content, with this qualification: whereas we have imagined that *M*'s belief is related in a characteristic way to the light's having looked as it did to him, *N*'s belief is not related in the same way to the light's having looked in some way to him. Although *N*'s belief is identical in content to *M*'s, it differs from *M*'s in that it is not, as we have characterized it so far, in any way rational. We needn't suppose, however, that insentient perceptual beliefs such as *N*'s must be entirely without rationality; we can imagine perceptual beliefs unaccompanied by the perceived object's appearing in some way which are nevertheless supported by reasons. *N* might support his belief, for example, by an inductive argument appealing to the consideration that whenever he has been at this intersection before, he has experienced the inclination to believe that he has just acquired through the use of his eyes the information that there was a red light at the intersection, and further that in all such cases, the belief he has been so inclined to accept has upon subsequent investigation turned out to be true. Or he may have been told by someone whose judgment he accepts as authoritative in matters of this sort, perhaps a psychologist with whom he is cooperating as the subject in an experiment on non-conscious perception, that his (*N*'s) present inclination to believe that there was a red traffic light at the intersection derives from his just having subliminally seen a red traffic light. I will refer to these two possible kinds of reasons for *N*'s adopting his perceptual belief as "induction" and "testimony," respectively. A different reason for his adopting that perceptual belief is the light's having looked in some way to him. I shall call this kind of reason "sensation." Now what goes for *N* also goes for *M*—there are several possible reasons for his adopting the belief that he has just acquired, through the use of his eyes, the information that there was a red traffic light at the intersection. As we have imagined his case *M*'s reason for his belief is sensation. This only means that, as we have already seen, *M*'s belief and his way of justifying it are intimately related, in a way yet to be explicated, to the way the light looked to him at t_1. Although sensation is the only sort of reason we find natural to imagine, under any readily conceivable conditions, as supportive of perceptual beliefs, it is nevertheless only one among other possible kinds capable of conferring degrees of rationality upon perceptual beliefs.

Recognizing that there are alternative reasons *M* could have for adopting the belief that he has, through the use of his eyes, acquired the information that there was a red traffic light at the intersection makes it clear that *M*'s statement "It looked red" can advance our understanding of that belief by identifying *M*'s reason for holding it, but without presenting that reason as an independent sufficient condition for the belief's rationality. For the explanation of his belief *M* provides by saying "It looked red" may now be compared to the explanation of a fever by attributing it to a deficiency of white corpuscles in the blood, but without identifying the deficiency as a sufficient condition for the fever. In such cases our explanation takes the form: "Given that the fever occurred, and given that it must have been caused either by a deficiency of white corpuscles or by an infection in the lymph system, or by some unknown condition, and given further that the conditions pretty much rule out the latter two possibilities, the fever must have been brought on by deficiency of white corpuscles." We do not identify the relevant disorder as the sufficient condition for the occurrence of the fever (since, for one thing, we haven't said what the degree of deficiency of white corpuscles was) but as the condition that was, under the circumstances, necessary for its occurrence. Similarly the increase in our understanding of his belief provided by *M*'s saying "It looked red" may now be seen to derive from our recognizing *M*'s reason as the one necessary for the rationality of his belief, given that it is a belief of that kind (a perceptual belief) and given that it is clear *M* has not accepted his belief on the basis of any unusual or unorthodox reason for perceptual belief such as testimony or induction.

Providing an explanation of this form is, I suggest, the primary force of *M*'s statement "It looked red." His statement increases our understanding of *M*'s belief by making it clear what kind of belief it is—identifying it as a perceptual belief—and at the same time articulating what his reason must be, the circumstances having ruled out such unusual reasons as testimony or induction.

While *M*'s explanation and the fever-explanation are comparable, there is this important difference between them: whereas in the case of the fever we have a conception of the explanans as an independent necessary condition or as an independent sufficient condition of what it explains (whether or not identifying it as such plays a role in the explanation we give), and are therefore not prohibited on conceptual grounds from speaking of its existing independently of what it explains, we have no conception—other than the notoriously problematic one already rejected—of the relevant sensation—the light's looking red to *M*—as an independent condition for the rationality of *M*'s belief, and accordingly we are not free to speak of its existing independently of what it explains.

I can imagine someone's questioning whether reference to explanatory sensation can provide genuine explanation of perceptual belief. In order for a traffic light to have looked red to me at time t_1, where its having looked red to me is an explanatory visual sensation, it must now be possible for me to explain the relevant belief (in part) by reference to the light's having looked red to me. An explanatory sensation is introduced, that is, specifically with reference to the belief it is supposed to explain. Won't explanations afforded by explanatory sensation then always be vacuous in the sense of failing to add any real information to what is said in the description of what they explain—like explanations of the sleep-inducing property in terms of soporific powers?

This objection would hold only if an explanatory sensation could be defined exclusively in terms of the belief it explains. But it can't; beyond the belief that a given explanatory sensation explains, there are further, equally essential conditions for the sensation, such as its relations to elements in the experience it does not purport to explain. For example, an explanatory sensation consisting in a thing's having looked red to me will have necessary causal conditions in the physical stimuli and physiological states underlying the experience, which conditions cannot be defined in terms of the problematic belief. Again, an explanatory visual sensation consisting in a thing's having looked red to me could harmonize with, but could not severely clash with, a concurrent visual sensation consisting in a thing's looking yellow to me. The relation of an explanatory sensation to the belief it explains, then, while more than empirical, is not so strong as to make it true by definition that if the latter has occurred, then the former has also. Hence reference to explanatory sensation can carry genuine explanatory force.

IV

Interpreting the explanation provided by *M*'s statement as explanation by identification of a dependent necessary condition for the rationality of his belief makes possible our interpreting the sensation in *M*'s seeing as a sensation that is explanatory in the way that the thought that there was another stair can be an explanatory thought. For what makes the thought that there was another stair an "explanatory" thought is that it is not independent of the speaker's stumbling and being able to explain his stumbling by saying "I thought there was another stair." This thought is not independent of the whole history of the episode of which it is a part, and so does not explain the man's stumbling by being an independent condition for his stumbling. Rather, this thought is what must have brought about the stumble—given that the stumble was not deliberate and not attributable to an obstacle on the stair or to darkness, and so on. And just as one's ability to explain in this way why he stumbled is what it is for him to have had the thought—an explanatory thought—that there was another stair at the time of his stumbling, so *M*'s ability to explain his current perceptual belief that the light was red by saying "It looked red" is what it is for him to have had an explanatory visual sensation at the time of his seeing.

Notice that the explication of *M*'s statement I have offered cannot be formulated along the lines required for the "independent sufficient condition" explication; it is quite different in form from "If testimony is received, inductive evidence is received, or sensation occurs, they will justify the belief that there was a red traffic light at the intersection." First, my explication states necessary, while the other states sufficient conditions. Secondly, the explication I am rejecting requires a comparatively more specific characterization of the relevant sensation. As long as we hold to the "independent sufficient condition" model of the explanation effected by *M*'s statement "It looked red," it would seem necessary for the unpacking of that statement to be relatively specific—to indicate, for example, roughly how long the thing looked red to *M*, and roughly how large and how bright it looked. For surely no defensible "rule of evidence" would sanction our holding, under a given set of independent background beliefs, that just any independently existing sensation consisting in a thing's looking red and round to *M* would be sufficient to justify *M*'s believing that there was a red traffic light at the intersection. On the other hand, no corresponding specificity is required when *M*'s explanation is explicated in the way I have suggested, any more than the explanation of stumbling provided by an explanatory thought requires us to be able to say quite precisely what the relevant thought was like. In each of these cases, since the entire meaning of the relevant statement ("I thought there was another stair," "It looked red") is explicated solely in terms of the role which these statements have in the explanation of related psychological states, there is no need, and no justification, for supposing the thought or the visual sensation in question to possess determinate characteristics beyond those explicitly entering into the explanations these statements provide.

Part of what this result of the present analysis comes to, then, is that we have a coherent way of thinking about the possible indeterminacy of the sensory content featured in non-conscious seeing. If, for example, *M* had non-consciously but rationally acquired the perceptual belief that there had been a striped tiger standing at the intersection, the question could arise "Did the striped tiger's looking striped to *M* include its looking to have a determinate number of stripes?" If *M*'s experience was such that we want to say "No," i.e., that though the tiger looked to *M* to have yellow and black stripes it did not look to him to have any determinate number of stripes, we have a way of justifying this. For we can say that the tiger's looking striped to *M* included his looking to have exactly thirteen stripes if and only if *M* can further explain what he now believes by saying "It looked to have exactly thirteen stripes." And the tiger's looking striped to *M* did not include his looking to have a determinate number of stripes if and only if nothing *M* says or needs to say in explanation of his belief includes his referring to a determinate number of stripes. In this case there is no such thing as the tiger's having looked to have just so many stripes, any more than there is such a thing as an explanatory thought's having a precise character. Not that the tiger might have looked to *M* to have thirteen stripes without *M*'s noticing it. Rather, its looking to *M* to have thirteen stripes is, simply, a matter of what enters into explaining *M*'s current perceptual belief.

V

It will be objected that it is not possible for explanatory sensation and subsequent belief to be related in explanation unless they are related in fact, and if they are related in fact then there *was* an earlier sensation.

This objection derives from a misconception of the kind of explanation an explanatory sensation provides. An explanatory sensation "explains" a perceptual belief, not as an hypothesized cause of that belief, but by contributing to a fuller, more adequate description of that belief. It is in order to say something about a perceptual belief, in the sense of further articulating it, making clear its character as a rational perceptual belief, that one mentions the explanatory sensation. The condition of an explanatory sensation's occurring is not that it was a discrete event which occurred prior to a belief as its partial cause, but rather that we are now able to further describe or elaborate upon the belief by reference to the sensation. And this is what we in fact do; we engage in this way of elaborating upon, commenting further upon the perceptual beliefs we acquire in the kind of non-conscious perceiving we have been discussing by referring to prior visual sensation, and in doing so we do not seem to mean that the prior sensory elements cited are causes of the belief we are trying to explain. (This does not in itself mean that the explanatory sensation and other psychological states cited in the explanation of one's perceptual belief cannot be (partial) causes of the beliefs they explain. Explanation by redescription can be causal, as when we explain why a man went to the pantry by saying that he went because he wanted to get some cheese. What it means, rather, is that it is such redescription of perceptual belief, not identification of its causes, that contains the nerve of explanation provided by appeal to explanatory sensation.)

At this point the objection that an explanatory sensation and subsequent belief cannot be related in explanation unless they are related in fact may persist, as follows: "How can we 'further describe or elaborate' by reference to a sensation that did not really occur?" In this form, the objection misses the crucial point of what I am claiming. What I am claiming does not include a denial that there was a sensation. Of course, if we were to accept that doctrine of the mental according to which one's having had a visual sensation in a situation must consist in there having been a temporally discrete mental occurrence, then one's being able to say, in explanation of a belief he has adopted, "It looked red" would not count as his having had a visual sensation in that situation. Thus by comparison with that doctrine, I shall appear to be denying that *M* had a visual sensation. But I have given reasons for questioning that doctrine of the mental in Sect. II, above. As with explanatory thought, we are entitled to question the presumption that M either had the sensation in question, or else didn't, independently of what comes after. This presumption is foreign to our ordinary way of thinking about such cases. The unphilosophical answer to the question whether M really had a visual sensation is not categorical. To the extent that we accept the remark "It looked red to M" as adding something essential to our understanding of M's believing that there was a red traffic light at the intersection, to that extent our ordinary way of thinking about such cases commits us to asserting the existence of the relevant sensation. And in the event that that explanation were somehow found to add nothing to our understanding of M's believing that there was a red traffic light, our ordinary way of thinking would condone our relinquishing the assertion that there was a sensation. Thus according to our ordinary way of thinking the entire meaning of the existence-claim "M had a visual sensation" derives from the role which that claim has in the subsequent explanation of his belief. And what I am claiming is that M *did* have a visual sensation at t_1, in that he can now, at t_2, add something essential to our understanding of his perceptual belief that there was a red traffic light at the intersection, by saying "It looked red."

To sum up: the foregoing account of the sensory element in non-conscious seeing takes seriously the similarity of reports of the sensory element in non-conscious seeing to the characteristically explanatory remark "I thought there was another stair." This account accords well with the *de facto* explanatory character of such reports and with the most natural explication of the kind of explanation such reports give. In addition the analysis enables us to make sense of the possible indeterminacy of the sensation featured in non-conscious seeing. Finally, I have argued in Sect. I, above, that an adequate analysis of non-conscious seeing must make it possible for us to acknowledge the sensory component of non-conscious seeing while rejecting both the notion of a sensory event that takes place unnoticed in the percipient's mind, later surfacing to conscious memory, and the reduction of non-conscious seeing to mere physiological states. The explanatory sensation analysis allows us to do that.[1]

Colorado State University

Received September 9, 1971

---

[1] I am indebted to Donald Gustafson, Berel Lang, and Moreland Perkins who are not responsible for the result.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 197

VI. LOCKE ON THE IDENTITY OF PERSONS

BARUCH BRODY

LOCKE'S discussion of personal identity is justly famous. It has not often been noted, however, that it is found within the context of a much more general theory about identity, a theory which (if true) raises considerable doubt about the point of his discussion of personal identity. I would like, in this paper, to analyze the relationship between his general theory of identity and his theory of personal identity. My purpose is twofold. On the one hand, I want to show that, given Locke's general theory, a theory of personal identity of the type that he advances faces much more fundamental problems than the familiar ones raised by Reid and Butler. On the other hand, I want to show that this is not a problem uniquely faced by Locke; other theories face it equally. I have argued elsewhere[1] that any theory of personal identity (or of the identity of any other type of substance) faces such a problem. What I want to show here is how concretely such problems arise, and I am using Locke as my concrete example of the problems arising.

I

Locke, when he first states his general account of identity, offers two different versions of it. The first is:

> That, therefore, that had one beginning, is the same thing; and that which had a different beginning in time and place from that, is not the same but diverse. (*Essay Concerning Human Understanding*, II, 27, 1.)

The second is:

> From what has been said, it is easy to discover what is so much inquired after, the *principium individuationis:* and that, it is plain, is existence itself; which determines a being of any sort to a particular time and place, incommunicable to two beings of the same kind. (*ibid.*, II, 27, 4).

What do these two accounts come to? The first seems to offer the following claim: (1) x and y are identical if there is some moment in time t_1 and some place p_1 which are such that t_1 is the first moment at which x existed and at which y existed and they were both in p_1 at t_1 while x and y are different if they both began existing but there is no t_1 and p_1 which jointly satisfy that condition. The second offers the following claim: (2) x and y are identical if they are of the same kind and, for any t and p, x is in p at t if and only if y is.[2]

There are, obviously, several important differences between these two accounts. To begin with, while both offer a sufficient condition for identity, only the first offers a sufficient condition for diversity and, consequently, a necessary condition for identity (viz., the nonsatisfaction of the sufficient condition for diversity). Secondly, the first account places special significance upon where an entity is at the moment at which it begins to exist while the second account sees its location at any moment as being of equal significance. Finally, while the first account deals only with the identity of entities that came into existence at some time, the second account applies more generally to any entities that have spatio-temporal location even if they always existed.

What lies behind these two accounts of identity? If we look at the very beginning of Locke's discussion, we find that he developed the following argument for (1):[3]

> . . . we never finding, nor conceiving it possible, that two things of the same kind should exist in the same place at the same time, we rightly conclude, that, whatever exists anywhere at any time, excludes all of the same kind, and is there itself alone. When therefore we demand whether anything be the same or no, it refers always to something that existed such a time in such a place, which it was certain, at that instant,

---

[1] In "Is There a Philosophical Problem about the Identity of Substances," a paper read to the western division meetings of the American Philosophical Association in 1969 and published in *Philosophia*, vol. 1 (1971), pp. 43-60.

[2] An alternative interpretation would only require, for the identity of x and y, that there be some p and t which are such that x and y are both in that p at that t. This claim collapses, however, into our (2) unless you allow for the possibility that the same entity might be in two different places at the same time, a possibility that Locke explicitly rules out (in II, 27, 1).

[3] Unfortunately, he offers no explicit argument for (2), claiming merely that it follows from what he has already said.

> was the same with itself, and no other. From whence it follows, that one thing cannot have two beginnings of existence nor two things one beginning; it being impossible for two things of the same kind to be or to exist in the same instant in the very same place; or one and the same thing in different places. (*ibid.*, II, 27, 1.)

The argument seems to run as follows: (a) if x is in p_1 at t_1 and y is of the same kind as x and $y \neq x$, then y is not in p_1 at t_1; therefore, (b) two things cannot have one beginning of existence and (c) one thing cannot have two beginnings of existence; consequently, (d) principle (1) of identity is true.

There are two things that should be noted about this argument. The first is the very strange claim that (c), with its consequential sufficient condition for diversity, the one stated in (1), follows from (a). This is obviously a mistake. What Locke needs to get (c) is some additional principle like (a*) if x is in p_1 at t_1 and $p_1 \neq p_2$ then x is not in p_2 at t_1. Locke seems to recognize this when, in the last clause of the above-quoted argument, he adds such a principle. But, as the earlier parts of that quotation indicate, Locke normally did not see this point and thought, quite incorrectly, that (c) follows from (a). The second point to be noted is that even (b) doesn't follow from (a). After all, it only follows from (a) that two things *of the same kind* cannot have the same beginning.

If, then, we were to correct Locke's argument to avoid these problems, we would wind up with the following argument: (a′) if x is in p_1 at t_1 and y is of the same kind as x and $y \neq x$, then y is not in p_1 at t_1; therefore (b′) two things of the same kind cannot have the same beginning of existence (cannot come into existence in the same place at the same time); consequently, (c′) x and y are identical if they are of the same kind and if there is some moment in time t_1 and some place p_1 which are such that t_1 is the first moment at which x existed and at which y existed and they were both in p_1 at t_1 (we shall call this principle of identity (1′)).

Now Locke's principle (2) also follows from (a′). It is, moreover, a better principle of (1′). partially because it applies to entities that have always existed, and not merely to entities that have come into existence, and partially because it recognizes no special significance for the moment at which the entities come into existence, a moment that, in light of (a′) (the rationale for both of these principles), has no special importance. It is not surprising, therefore, that when Locke finally stated his principium individuationis, he dropped (1) — and (1′) which, presumably, was what he had in mind — and used (2) instead. We shall, for the rest of this paper, follow his lead and treat (2) as Locke's general sufficient condition for identity.

Principle (2) raises several important issues. To begin with, it only states a sufficient condition for the identity of x's and y's which are spatio-temporally locatable. What about entities that are not so locatable? Locke actually considers one entity which might meet this description, God, but he has the following to say about God's identity:

> God is without beginning, eternal, unalterable, and everywhere, and therefore concerning his identity there can be no doubt. (*ibid.*, II, 27, 2.)

In other words, Locke sees God as literally existing at all times and at all places and, therefore, feels that his standard theory of identity can be applied to questions about God's identity and that, given his unique features, such problems are easily solvable. Another example that one would have hoped that Locke would have considered is that of numbers, but Locke does not, presumably because he thinks of numbers as modes and not as substances. In short, then, Locke probably did not see this as a problem because he did not see that there were any substances that are not locatable in space and time.[4]

The second issue that arises in connection with Locke's principle (2) is the meaning of "the same kind." Which kinds are relevant here? Locke offers a clear answer to this question, viz., that the three relevant kinds of things are God, finite spirits, and particles of matter.[5] In other words, Locke is saying that x and y are identical if they are in the same places at the same times and have certain properties (the ones that make them God, or a finite spirit, or a particle of matter) in common. The question that naturally arises is: What makes these properties special? What reason is there to suppose that the common possession of these properties, and not others, makes it impossible for two objects to be in the same place at the same time?

---

[4] There is no indication as to what he thought about angels. If they exist, they would be substances and (presumably) would not be in space and time. The medievals recognized, of course, that there were serious problems with the individuation of angels.

[5] The relevant text here (II, 27, 2) offers no support for the suggestion made by David Sanford ("Locke, Leibniz, and Wiggins on Being in the Same Place at the Same Time," *Philosophical Review*, vol. 79 [1970], pp. 75–82) that Locke means his thesis to be applied primarily to material things. The passage in I, 1, 18 that Sanford cites, where Locke states his principle only for material things, hardly supports his view since he is only stating there lots of common maxims.

How might Locke avoid such problems? Well, if we think of Locke's theory as it now stands as running into trouble just because it picks out certain properties as special (in the sense that the common possession of them rules out being in the same place at the same time), then we are naturally led to suspect that such problems will be avoided if we treat all properties equally. This leaves open two possibilities, the theory that (2′) x and y are identical if they are in the same place at the same time and the theory that (2″) x and y are identical if they are in the same place at the same time and have all of their properties in common. (2′) faces, however, serious difficulty. After all, my body and I are in the same place at the same time, but they cannot be identical since there are some properties that they do not have in common. More generally, there can be two entities in the same place at the same time but which do not have all of their properties in common, so being in the same place at the same time cannot be a sufficient condition for identity and (2′) must be rejected. Now (2″) avoids this type of problem entirely. It would appear, therefore, that (2″), which imposes a more restrictive condition for identity than (2), is the most plausible version for a general theory of identity like Locke's, although, unfortunately, it is not the one that Locke actually adopted.[6]

II

So much for Locke's general theory of identity. We know, however, that this is not the whole of Locke's theory of identity, for he goes on to claim that

> to conceive and judge of it [identity] aright, we must consider what idea the word it is applied to stands for; it being one thing to be the same substance, another the same man, and a third the same person, if person, man, and substance are three names standing for three different ideas. (*ibid.*, II, 27, 8.)

In light of this claim, Locke goes on to develop special theories for the identity of men and of persons. We will return below to a consideration of these theories; what we have to consider first is what Locke meant by this claim and why, in the light of it, he developed these special theories.

One way to pose the problem is as follows: there is a trivial way in which what Locke says is true. For x and y to be the same person, x must be a person, y must be a person, and x must be identical with y. For x and y to be the same man, however, the first two of those requirements need not be met and, instead, x and y must both be men. The requirements, then, for being the same man and for being the same person are different. But all that this trivial truth means is that, in determining whether x and y are the same P, we have to determine whether they are both P's and whether they are identical, and how we do the former depends upon what P is. But in no way does it follow that what P is will affect the way we determine that x is identical with y. It is clear, however, that Locke does not have this trivial point in mind because, for different P's, he wants to offer different requirements for their identity. What Locke wants to claim, in other words, is that what P is will affect the way in which we determine whether x and y are identical. But why does Locke want to claim this? Why does he offer these special principles? He already has (in (2)) a general theory of identity equally applicable to all substances.

A good way to approach this question is from a consideration of Locke's distinction between a mass of particles and an oak tree. Locke says that

> If two or more atoms be joined together into the same mass . . . whilst they exist united together, the mass, consisting of the same atoms, must be the same mass, or the same body, let the parts be ever so differently jumbled. But if one of these atoms be taken away, or one new one added, it is no longer the same mass or the same body. In the state of living creatures, their identity depends not on a mass of the same particles, but on something else. For in them the variation of great parcels of matter alters not the identity: an oak growing from a plant to a great tree, and then lopped, is still the same oak. (*ibid.*, II, 27, 4.)

What Locke is saying here is that the identity of compound substances (substances composed out of other substances) is sometimes dependent upon the identity of the substances composing it and is sometimes not. In particular, while a mass of particles m_1 is identical with a mass m_2 if and only if the set of particles making up m_1 has the same membership as the set of particles making up m_2, an oak tree o_1 can be identical with an oak tree o_2 even though the set of particles making up o_1 is not identical with the set of particles making up o_2.[7] Or, (3) for all m_1 and m_2 which are masses of particles, $m_1=m_2 \equiv \{x|x \text{ are components of } m_1\} = \{x|x \text{ are components of } m_2\}$, but $\sim$(4) for all o_1 and o_2 which are oak trees, $o_1=o_2 \equiv \{x|x \text{ are components of } o_1\} = \{x|x \text{ are components of } o_2\}$. It should be noted that although Locke rejects (4), he offers in the next paragraph a principle for the identity of oak trees, the principle that

---

[6] The theory that I presented in the paper cited in the first footnote is, in effect, a combination of (2″) with a theory for the identity of substances that are not spatio-temporally locatable. The reader is referred to that paper for further details of the theory.

> That being then one plant which has such an organization of parts in one coherent body, partaking of one common life, it continues to be the same plant as long as it partakes of the same life, though that life be communicated to new particles of matter vitally united to the living plant, in a like continued organization conformable to that sort of plants. (*ibid.*, II, 27, 5.)

Roughly speaking, this principle is (5) for all o_1 and o_2 which are oak trees, $o_1=o_2 \equiv o_1$ and o_2 partake of the same life in a like organization.

There are two questions that arise about (3) and (5). One has to do with Locke's reasons for advancing these principles. An answer to this question will, presumably, be an answer to our original question as to why, given a theory of identity like (2), Locke felt that whether x and y are identical depends upon what type of thing x and y are. After all, it is just principles like (3) and (5), which introduce different identity conditions for different types of entities, which are required by Locke's original claim. The second question that we shall have to consider is whether principles like (3) and (5) can be conjoined with (2) without producing serious difficulties in the theory of identity.

The least charitable interpretation of Locke is that he has fallen into a pseudo-problem about how x (at time t_1) can be identical with y (at time t_2)[8] if the objects making up x are not identical with the objects making up y and that all of the conditions offered in the principles like (5) are meant to substitute for the satisfaction of the condition of having the same constituents. I say that this is a pseudo-problem because, on the whole, there is no connection between x's being identical with y and their composition. Moreover, Locke recognized all of this; after all, with the exception of masses of particles, he explicitly claimed that x and y could be identical even if they do not have the same constituents. So we should look for some other motivation for principles like (3) and (5).

The following account seems correct: Locke was committed to the view that problems about identity arise only when we are considering an entity which we know to exist at t_1 and wondering whether or not it is identical with some entity that we know to exist at a later t_2. As he said:

> Another occasion the mind often takes of comparing, is the very being of things, when, considering anything *as existing at any determined time and place*, we compare it with itself *existing at another time*, and thereon form the ideas of identity and diversity. (*ibid.*, II, 27, 1. The italics are Locke's.)

Thinking about identity problems this way,[9] one might easily wonder as to how different y at t_2 can be from x at t_1 and still be identical with that x, i.e., one might easily wonder as to what changes an object can undergo while still remaining the same object. Principles like (3) and (5) offer, in effect, answers to such questions for different types of entities. (3) says about masses of particles that they cannot gain or lose any members; if they do, the old mass no longer exists and there is a new one present. (5) says about oak trees that they must retain certain basic organizational features and continue to participate in the same life. In other words, Locke noticed that principles like (2) do not do justice to all of our ordinary intuitions about identity. It is ordinarily thought that in order for x at t_1 to be identical with y at t_2, y at t_2 must retain some of the properties that x had at t_1. Or, to put it another way, if y at t_2 is a radically different type of entity than x at t_1, they cannot be identical. Now none of these intuitions are captured by (2), for all that (2) requires is that x and y both be physical objects, or finite spirits, or God and that they both be in the same places at the same time, i.e., that x be at t_2 where y is at t_2 and that y was at t_1 where x was at t_1. As far as (2) is concerned, y could have been a door while x

---

[7] This does not contravene Leibniz' Law since o_1 would be an oak tree at time t_1 while o_2 would be an oak tree at a different time t_2.

[8] Of course x at t_1 and y at t_1 must have the same constituents or they are not identical. But that is required by Leibniz' Law and not by some special principle.

[9] And therefore disregarding cases such as when one wonders whether Elizabeth II is identical with the present Queen of England, knowing very well that they both exist now but not remembering what country she is queen of.

was an oak tree and yet x be identical with y. It is a principle like (5) that rules such a possibility out, and Locke adopted principles like (3) and (5) to capture this element in our ordinary intuitions about identity.

It should be noted that the adoption of such principles gives rise to an interesting sort of essentialism. We shall say that x has P essentially just in case x has P, P is a non-tensed property, and if x were to lose P, then x would no longer exist. Roughly speaking, an object has a property essentially if it cannot lose that property without going out of existence.[10] The adoption of a principle like (2) only commits us to a very minimal form of essentialism since the only essential properties are being a physical object, or being a finite substance, or being a God. According to (2), retaining one of these properties is the only requirement of the retention of a non-tensed property that must be met for x to continue to exist at t_2, for there to be some entity at t_2 identical with x which existed at t_1. But the adoption of principles like (3) and (5) require x, in order to continue existing, in order to be identical with some entity existing at t_2, to retain additional non-tensed properties (the property of being composed out of certain other substances, or the property of having a certain kind of organization, or . . .) and commits us therefore to stronger essentialist claims.

In summary then, Locke felt that, for any given entity, there were only certain changes that it could undergo and still continue to exist, that it had to retain certain properties from t_1 to t_2 if there was going to be any entity existing at t_2 identical with it. Moreover, he felt that what properties these were would vary from one type of entity to another. In some cases, the properties in question were composition-properties; but usually this was not the case. All of this being so, whether x at t_1 is identical with y at t_2 depends upon what type of entity x and y are and upon what principles of identity (like (3) or (5)) govern the identity of such entities. Since (2) tells us none of this, it will not do as a complete theory of identity.

All of this raises, however a serious problem for Locke. After all, principle (2) says that it is sufficient for x and y (providing that they are both physical objects, or finite spirits, or God) being identical that they be in the same place at the same time. And, although this is not included in (2), we would certainly want to say, by virtue of Leibniz' Law, that their being in the same place at the same time is necessary for their identity. Now if we add some additional principle (like (3) and (5)) that state additional necessary and/or sufficient conditions, we are, in effect, making commitments whose justification is very unclear. For example, consider some mass of particles m_1 existing at t_1 in p_1 and some mass of particles m_2 existing at t_2 in p_2. Now Locke's principle (3) tells us that $m_1 = m_2$ if and only if the particles making up m_2 at t_2 are identical with the particles making up m_1 at t_1. But (2) tells us that $m_1 = m_2$ if and only if m_2 was in p_1 at t_1 and m_1 was in p_2 at t_2. So it follows that m_2 was in p_1 at t_1 and m_1 was in p_2 at t_2 if and only if the particles making up m_2 at t_2 are identical with the particles making up m_1 at t_1. It is difficult to see what force this latter claim has; it certainly seems that there could be counter-examples to it. So the fact that it follows from (2) and (3) suggests that we ought not jointly to adopt the two of them. More generally, by adopting principles like (3) and (5), we are committing ourselves to the view that some condition C is necessary and sufficient for x's (of a certain kind) being identical with y's (of the same kind). But, according to (2), $x = y$ if and only if they are in the same places at the same time (and are both physical objects, finite spirits, or God). It follows then that x and y can be in the same places at the same times if and only if C obtains. If, as often seems to be the case, we can imagine counter-examples to that last claim, are not we going to have to reject either (2) or Locke's additional principles?

It should be noted that this is not only a problem for Locke. Anyone who is prepared to accept the identity of indescernibles (our (a″)) and yet wants to say that physical object o_1 is identical with physical object o_2 only if they are spatio-temporally continuous has the same problem. And even if one is not prepared to accept the identity of indescernibles, one will still be in the same trouble if one adopts Leibniz' Law and also claims that spatio-temporal continuity is sufficient for the identity of physical objects. The problem, in its most general form, is a simple one. If you have (because of the adoption of the identity of indescernibles) a view that a certain condition C_1 is sufficient for identity, then the adoption of the view that some other condition C_2 is necessary for identity of a certain type of entity commits you to the claim that C_1 can be satisfied by entities of that type only when C_2 is. Similarly, if you have (because of Leibniz' Law) a view that a certain condition C_3 is necessary for identity, then the adoption of the view that some other condition C_4 is sufficient for identity of a certain type of entity commits you to the claim that C_4 can be satisfied by entities of that type only when C_3 is. Since these conditions often seem unrelated, and, consequently, these commitments often seem to be unjustified, this result casts doubt upon theories of identity that contain both of these conditions.

---

[10] I have discussed this type of essentialism more fully in my "Natural Kinds and Real Essences," *Journal of Philosophy*, vol. 64 (1967), pp. 431–446, "Two Recent Contributions to the Theory of Reference," *Studies in the History and Philosophy of Science* (1971), and "De Re and De Dicto Interpretations of Modal Logic," *Philosophia* (1972).

Intuitively, the situation is as follows: Locke (and many of us) seem committed to (2) or to some modification of it like (2″). But even if we are not, we are certainly committed to the view that being in the same places at the same time is a necessary condition for identity. This, after all, follows from Leibniz' Law. On the other hand, we want, like Locke, to impose additional necessary and sufficient conditions for the identity of entities of different types, conditions that are based upon our intuitive feelings that entities of those types cannot change in certain ways without thereby going out of existence. But when we add these additional conditions, we find ourselves committed in ways that are difficult to justify. Two options are therefore open to us. We can reject these additional conditions (and reject essentialism) by rejecting the claim that there are certain changes that an object can undergo which necessarily mean its going out of existence. Or, we can accept these conditions and the commitments that come along with them. As far as I can see, most writers on identity have followed Locke, both in accepting these commitments and in not realizing that they have done so. Whether they have been wise in doing so is a matter that lies beyond the scope of this paper.

III

In light of this understanding of Locke's general theory of identity, we turn to his discussion of personal identity. The thesis that he wanted to maintain is that

> When we see, hear, smell, taste, feel, mediate, or will anything, we know that we do so. Thus it is always as to our present sensations and perceptions: and by this everyone is to himself what he calls self. . . . For, since consciousness always accompanies thinking, and is that which makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking beings, in this alone consists personal identity, i.e., the sameness of a rational being: and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it that that action was done. (*ibid.*, II, 27, 11.)

From our perspective, Locke is adding here a new principle like (3) and (5), one which applies to persons. What is this new principle? Two candidates suggest themselves: (6) for all x and y which are persons, $x=y\equiv x$ at t_2 remembers experiencing and doing all that y at t_1 was conscious of experiencing and doing or (6′) for all x and y which are persons, $x=y\equiv x$ at t_2 remembers experiencing or doing something that y at t_1 was conscious of experiencing or doing. Which of these claims is the one advocated by Locke is something that we shall consider in a moment;[11] what we should note for now is that, on either interpretation, the principle is about the type of change that an object can undergo while still being the same object. x at t_2 can be identical with y which existed at t_1 only because there has not been such a change in the content of the minds in question that x is not conscious of what y was conscious.

It seems clear that it is (6′), and not (6), that Locke was advocating. To begin with, (6) is totally implausible since it guarantees that no person is identical with anything that existed just a few hours ago. After all, who now remembers every action and experience that he was conscious of doing and experiencing a few hours ago. So why attribute to Locke the totally unplausibly (6) when we have the much more plausible (6′) to attribute to him? Secondly, if we look at what Locke has to say about certain cases of non-identity, we see that it is (6′), and not (6), that he is operating with. Thus, in talking about the difference between the identity of persons and the identity of men, Locke says:

> But if it is possible for the same man to have distinct incommunicable consciousness at different times, it is past doubt the same man would at different times make different persons; which, we see, is the sense of

---

[11] Both of these interpretations have to be distinguished from A. Flew's interpretation (in "Locke and the Problem of Personal Identity," *Philosophy*, vol. 26 [1951]) pp. 53–68 that Locke is saying that (6″) $x=y\equiv x$ remembers at t_2 his doing what y did or felt at t_1. Flew's reading has no support in the text.

> mankind in the solemnest declaration of their opinions, human law, not punishing the madman for the sober man's actions, nor the sober man for what the madman did—thereby making them two persons. (*ibid.*, II, 27, 30.)

The sober man and the madman are not the same person because they have incommunicable consciousnesses, i.e., each is not conscious of anything the other is conscious of, and not merely because one does not remember some one thing of which the other is conscious. So it is (6′) and not (6), that Locke was working with.

It should be noted, in passing, that this means that a familiar objection to Locke's theory, Reid's gallant officer objection, is not, as it stands, correct. Reid[12] had argued as follows: (a) imagine that at time t_3 there is a general who remembers doing at t_2 an action (taking a standard) that was done by a certain gallant officer who, at t_2, was conscious of doing it. Moreover, imagine that at t_2 that officer remembers experiencing something at t_1 (being flogged) that happened then to a schoolboy who was, at t_1, conscious that it happened to him. Finally, imagine that, at t_3 the general does not remember that flogging; (b) by Locke's account of identity, the general at t_3 is identical with the officer at t_2 and that officer at t_2 is identical with the schoolboy at t_1; (c) therefore, by the transitivity of identity, the general is identical with the schoolboy; (d) however, according to Locke's theory, the general is not identical with the schoolboy because he does not remember being flogged; (e) therefore, Locke's theory is inconsistent and should be rejected. Now if Locke meant (6′), and not (6), step (d) of the argument should be rejected. For it is not necessary that the general remember the flogging in order to be identical with the schoolboy; he just has to remember something (in this case, something other than the flogging) which the schoolboy was conscious of at t_1. One can, however, modify Reid's objection to avoid this reply by stipulating that the general, at t_3, remembers nothing that the schoolboy was conscious of doing and experiencing at t_1. So modified, Reid's objection holds against (6′) as well as against (6).

Having before us (6′) as well as (5) and (3), we are now in a position to see that Locke has, in effect, adopted a very strange type of essentialism. It is necessary to distinguish two very different types of essentialism.[13] One, Aristotelean essentialism, is the view that every object belongs to some class (which could contain more than one member) which is such that all members of this class, but not any other objects, have a given property (or properties) essentially. The other is the thesis of individuating essentialism, the view that each object has a set of properties which only it has and it has them essentially. Now, it would seem that Locke is adopting this strange, very strong, second type of essentialism. After all, that common consciousness which a person at t_2 and a person at t_1 must have if they are to be the same person, that consciousness whose continued existence is necessary for the continued existence of the person and which is therefore essential to that person, is something that only one entity has.

This point can also be put as follows: Locke had adopted a series of principles that state what changes an object can undergo and still remain the same object. Masses of particles cannot gain or lose particles, oak trees must retain their organizational structure and their "common life," and people must retain some part of their consciousness. At first glance, it seems that these principles have the same import for each member of the class in question, but, on further analysis, this turns out to be a mistake. For example, what I have to retain to continue to exist is different from what you have to retain to continue to exist. I must remember some of the things that I am now conscious of experiencing or doing while you must remember some of the things that you are now conscious of experiencing or doing. Similarly, different masses must retain different compositions and different oak trees must continue to participate in different common lives.

The question that we have to consider therefore is very simple. We know that Lockie adopted these principles because he felt that (2) did not do justice to our intuitive feeling that objects do not continue to exist through radical changes. But to meet that intuition, it would be sufficient to adopt principles of Aristotelean essentialism, principles like (6*) if x at t_1 is a person, then y at t_2 is identical with x only if y is a person or (5*) if x at t_1 is an oak tree, then y at t_2 is identical with x only if y is an oak tree. Why then did Locke adopt this stronger form of essentialism?

---

[12] In his *Essays on the Intellectual Powers of Man* (Massachusetts, 1969), pp. 357–358.

[13] This distinction is a modification of the distinction drawn by R. B. Marcus in "Essential Attribution," *Journal of Philosophy*, vol. 68 (1971), pp. 187–202. In her definition of individuating essentialism, she does not require that the object have those properties essentially.

There is nothing in the text that fully answers this question, but the following suggestion seems plausible. Locke felt that what was needed was a special requirement for the identity of each object to distinguish different objects having the same Aristotelean essence. After all, given any class of objects having the same Aristotelean essence, the members of that class are not necessarily identical with each other, so any theory of identity must give you principles that distinguish them.[14]

If, however, this suggestion is correct, then it is evident that Locke has made a serious blunder here. After all, he already has a principle that enables him to distinguish objects with the same Aristotelean essence, viz., (2). If, then, like Locke, one is adding essentialist principles to a theory of identity that contains some principle like the identity of indescernibles, there is no need to go beyond Aristotelean essentialism.

IV

The following seems to have emerged in our consideration of Locke's theory of personal identity from the perspective of his more general views about identity:

(a) Locke had a general theory about identity which seems to be applicable to all substances.

(b) Nevertheless, he introduced special principles governing the identity of different types of objects because he felt that his general theory did not do justice to our intuitions about the type of changes that an object can undergo while still continuing to exist. Unfortunately, he did not recognize that this committed him to a variety of claims whose justification is very unclear.

(c) Moreover, the principles that Locke advanced are stronger than necessary to meet the intuitions in question and this additional strength seems unjustified.

If these claims are true, then Locke's theory of personal identity suffers from fundamental defects, much more fundamental than the familiar ones first raised by Butler and Reid. And if (2) or (2″) are plausible, then similar things can be said about all special theories of the identity of physical objects, persons, etc.

Massachusetts Institute
of Technology

Received June 25, 1971

---

[14] Principles like (6) do that because they state necessary *and* sufficient conditions; principles like (6*) do not because they only state necessary conditions.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 1972

VII. LEGITIMATE AUTHORITY

ROBERT F. LADENSON

TRADITIONALLY, a widespread belief among political philosophers has been that the primary philosophical issue concerning political authority is generated by a tension between the idea of morally legitimate political authority and the principle that one ought to follow the dictates of his conscience. This tension, they have felt, is best expressed by posing the question "Why is it that the members of the body politic have a duty to obey the commands of their rulers even on occasions when they oppose those commands in conscience?" In the first part of this paper I discuss the account of legitimate political authority advanced by Robert Paul Wolff, perhaps the most recent example of a treatment of the subject which takes the above question as its starting point.[1] The upshot of this discussion is that the question is not a genuine one; that is, that the conflict between morally legitimate political authority and the principle that one ought to follow the dictates of one's conscience does not exist. In the second part I set out what seems to me to be the outlines of a correct analysis of the concept of morally legitimate political authority and offer a few suggestions as to why so many philosophers have fallen into confusion on this matter.

I

In the preface to his recent monograph, *In Defense of Anarchism*, Robert Paul Wolff offers us the following interesting piece of intellectual autobiography:

> This essay on the foundations of the authority of the state marks a stage in the development of my concern with the problems of political authority and moral autonomy . . . during my first year as a member of the Columbia University Philosophy department I taught a course on political philosophy in which I boldly announced that I would formulate and then solve the fundamental problem of political philosophy. I had no trouble in formulating the problem—roughly speaking, how the moral autonomy of the individual can be made compatible with the legitimate authority of the state. I also had no trouble in refuting a number of supposed solutions which had been put forward by various theorists of the democratic state. But midway through the semester, I was forced to go before my class, crestfallen, and very embarrassed, to announce that I had failed to discover the grand solution.[2]

Wolff goes on to relate both how, after subsequent years of philosophical reflection, the problem remains for him as intractable as ever and how, indeed, he has become sceptical about whether a solution can be found. More will be said shortly about Wolff's scepticism. The point to focus upon now is that from the above passage it appears that Wolff believes the philosophical problem concerning political authority to arise out of an apparent or *prima facie* incompatibility between "the legitimate authority of the state" and the "moral autonomy of the individual." It will be useful then to dwell for a moment on the way in which Wolff explains these two key ideas.[3]

According to Wolff, all political societies (cities, nations, etc.) claim authority; that is, they claim the right to rule or, as Wolff frequently puts it, the right to command. To say of a given political society that its authority is *legitimate* is to say that its claims of authority are morally valid. Legitimate authority, for Wolff, is to be sharply distinguished from *de facto* authority. All that need occur for a state to have *de facto* authority is that many people believe its claims of authority to be morally valid. It need not be the case that they actually are. Throughout his monograph, Wolff uses the phrases "legitimate authority," "*de jure* authority," "the right to rule," and "the right to command" interchangeably.

---

[1] Among writers who clearly pose the philosophical problem of political authority in this way one finds Yves Simon, *Philosophy of Democratic Government* (Chicago, 1951), ch. 3; Nicholas Berdayev, "Slavery and Freedom" in *Patterns of Anarchy*, (ed. by Leonard Krimmerman and Lewis Perry, New York, 1965), pp. 150–165; Stephen Pearl Andrews, "The Science of Society" (also in *Patterns of Anarchy*, pp. 207–223); and Mulford Q. Sibley, *The Obligation to Disobey the Law* (The Council on Religion and International Affairs, 1970).

Both Rousseau (*The Social Contract*) and Jacques Maritain (*Man and the State*) also appear to adopt this formulation of the issue, although it is less clear in their cases that they do.

[2] R. P. Wolff, *In Defense of Anarchism* (New York, 1970), p. vii.

[3] Wolff's analyses of both legitimate political authority and moral autonomy appear in Part I of *In Defense of Anarchism*, *ibid.*, p. 3–18.

How does Wolff explain the other notion, that of moral autonomy? He begins by noting that while a man is reponsible for his actions whether he likes it or not, it is always up to him whether to take responsibility for them. Now one takes responsibility for his actions, on Wolff's view, when he both regards himself as the final judge about right and wrong with regard to his own conduct and when he acts accordingly; that is, when he deliberates to the best of his ability about what, in given circumstances, is the right thing to do. Thus one takes responsibility for one's own actions when he gathers all the information relevant for making a sensible decision on matters of right and wrong, carefully maps out the various possible courses of action, ascertains the reasons for and against each of them, and weighs those reasons against one another so as to decide what to do. Taking responsibility for one's actions in this sense is what Wolff means by moral autonomy.

There are then two key features of moral autonomy as Wolff conceives it. First, the morally autonomous person always attempts to the best of his ability to form a true opinion about right and wrong with respect to his own actions. Secondly, he is always prepared to act upon this opinion; that is, to do what he believes is right. To do the first without the second would simply be to engage in a form of sterile moral intellectualism. To do the second without the first (to do what one thinks is right without thinking the matter through carefully) would be to indulge in a form of well intentioned irresponsibility. In effect, Wolff's account of moral autonomy is a good explication of what is meant (or ought to be meant) by the claim that one ought to follow the dictates of one's conscience.

It will be recalled that according to Wolff, the fundamental problem of political philosophy is that of how "the moral autonomy of the individual can be made compatible with the legitimate authority of the state." It can now be seen that in saying this he is asserting that the problem to be faced is that of reconciling the principle that one ought to follow the dictates of one's conscience, as explained above, with the acknowledgment that at least in some cases, the state's claims of authority are morally valid. Furthermore, Wolff often formulates this task of reconciliation as the question of why it is that the citizen has a duty to obey the commands of his rulers simply because they are the commands of his rulers; in other words, even when he opposes those commands in conscience. In a long paragraph Wolff carefully distinguishes doing what someone tells you to do because you have independently decided that it is right, from doing what he tells you to do simply because he tells you to do it. Wolff then immediately goes on to say,

> The point of the last paragraph cannot be too strongly stressed. Obedience is not a matter of doing what someone tells you to do. It is a matter of doing what he tells you to do because he tells you to do it. Legitimate authority or *de jure* authority thus concerns the grounds and sources of moral obligation.[4]

Similarly, at other places in the monograph, Wolff poses the problem of political authority as being that of explaining why it is that one has a duty to obey the law simply because it is the law.[5]

The entire second and third parts of Wolff's monograph are devoted to various attempts at answering this question; that is, resolving the alleged tension between the idea of morally legitimate authority and the principle that one ought to follow the dictates of his conscience. Wolff examines various procedures for arriving at collective decisions and concludes that all except unanimous majority vote, which is impossible from a practical standpoint, fail to preserve the moral autonomy of the citizen. Thus, for Wolff, the conflict between the idea that there is legitimate political authority and the idea of the moral autonomy of the individual cannot be reconciled. Moreover, since he regards the latter as far more well entrenched than the former, he concludes that the former has to go. In other words he argues that the concept of morally legitimate political authority is incoherent. He tells us that no state has ever had or ever could have a moral right to rule. In this connection, Wolff compares his attack on the notion of morally legitimate political authority with Kant's destruction of speculative metaphysics in the first *Critique*.[6]

---

[4] *Ibid.*, p. 9.

[5] *Ibid.*, p. 9.

[6] R. P. Wolff, "On Violence," *The Journal of Philosophy*, vol. 66 (1969), pp. 601–616.

The boldness of Wolff's conclusions notwithstanding, surely something has gone wrong here. The belief that a state's claims of authority can be morally valid appears on its face to be utterly indisputable. Consider the familiar set of purposes enumerated in the Preamble to the United States Constitution; namely, "to establish justice, insure domestic tranquility, provide for the common defense, promote the general welfare, and secure the blessings of liberty to ourselves and our posterity." No one could reasonably deny that if there were a state which really did all that, its authority would surely be legitimate in a moral sense. Wolff's claim that the idea of morally legitimate political authority is incoherent implies that even after it has been scrupulously established that a given state does a good job of serving all of the above purposes, it still does not follow that the authority of that state is morally legitimate. Actually, an analogy more appropriate than the one drawn by Wolff would be between his doubts about the existence of morally legitimate political authority and the radical doubts about the existence of an external world which Kant was also concerned to refute. The parallel with radical scepticism in the realm of epistemology is almost exact. According to this kind of scepticism, even after we have checked all the considerations which, even in the widest sense, count as evidence for empirical knowledge, it still does not follow that the existence claims one makes about physical objects, presumably on the basis of this evidence, are well founded. If one asks the sceptic what would be needed in order to have good grounds for such claims, over and above the normal evidence of sight, touch, and so forth, he tells us that nothing would or even could do the job. Likewise with Wolff, if asked what would be needed to establish the moral legitimacy of a given state's authority, over and above showing that it does a tolerably good job of serving purposes such as those enumerated in the Preamble to the United States Constitution (i.e., the sorts of things one would naturally think of as relevant for deciding the issue) he would say that nothing could do this. On his view, no state, in principle has had or could have the right to rule. Now the beliefs which the radical sceptic about the external world seeks to undermine are so basic that they cannot be rejected. Accordingly, the fact that a given epistemological theory leads to radical scepticism about the external world is a sufficient reason for rejecting it. The same is true of Wolff's account of legitimate political authority. The fact that it leads to the kind of scepticism that it does shows his account to be unacceptable.

Wolff's views cannot be correct then but where do they go wrong? To answer this it is useful to focus again more carefully upon his problem so as to have a clear understanding of what is at issue. The question Wolff raises is that of why the citizen has a duty to obey the commands of his rulers (i.e., the laws) even on occasions when he opposes those commands in conscience. Wolff feels that to answer this question is to explain the basis of legitimate political authority; that is, to explain why it is that the state's claims of authority are, except perhaps in certain extreme circumstances, morally valid. He also feels that this question reflects a *prima facie* incompatibility between the belief that a state's authority can be morally legitimate and the belief that one ought to follow the dictates of one's conscience.

Now the question, "Why does the citizen have a duty to obey the law even when he opposes it in conscience?" is ambiguous. In asking this question one's intention could be to raise the issue of why morality frequently requires a person to obey laws which he believes to be unjust, or in some other way morally bad. That is, the aim of the question could be to ask on what grounds a person might correctly decide that though a given law to which he is subject is morally objectionable, it is nonetheless morally incumbent upon him to obey it.

This issue, though certainly important, does not seem to be the one that Wolff and others have tried to put their fingers upon in their discussion of legitimate political authority. The issue which interests them presumably involves the reconciliation of a belief in the legitimate authority of the state with a belief in the primacy of the dictates of one's conscience. This, however, does not at all figure in the question of why one is, at times, morally required to obey bad laws. A person might believe a law to be morally bad and yet, in spite of that, come to the conclusion that it would be wrong for him to disobey it. He might, for example, believe that though the law is bad, the institutions in the name of which it was decreed are good ones, and thus, deserving of the support that one characteristically shows for an institution by obeying the rules laid down by it. Even if the laws and legal institutions of his state are not very admirable a person may feel that it is morally incumbent upon him to obey because by violating the law he would undertake a substantial risk of being deprived of his freedom, and thereby be unable to look after people to whom he has responsibilities. In both of these examples the individuals involved would simply be choosing what they regard as the lesser of two evils. Consequently, in their respective cases compliance with morally bad laws is consistent with the principle that one ought to follow the dictates of one's conscience. There is no necessary conflict here.

One might, however, construe the question, "Why does the citizen have a duty to obey the law even when he opposes it in conscience?" in another way which seems to be closer to Wolff's concerns. Given that he perceives a conflict between morally legitimate political authority and the principle that one ought to follow the dictates of one's conscience, it would seem that the issue which lies behind his statement of the problem of political authority should be expressed thus:

> Why is it that a person should regard himself as morally required to obey the law even on those occasions when *all things considered* he believes that obedience would be morally wrong?

This formulation has the merit of focusing attention upon what seems to be the source of Wolff's anxiety. At the same time, however, it enables us to see more clearly that his anxiety is unfounded; that is, that it rests upon intellectual confusion. Anyone with an understanding of the concept of morality could not possibly consider himself as morally required to obey the law even when he believes that, *all things considered*, it would be wrong to do so. After all, what is it to make decisions about one's personal conduct from the standpoint of morality or, using Baier's phrase, to take the moral point of view? Whatever else it involves, it is to ascertain to the best of one's ability what is right and wrong, all things considered, with respect to one's own behavior and then to act accordingly. To claim that there are times when morality might require us not to do this is to betray one's misunderstanding of the concept of morality. To state the matter in another way, following the dictates of one's conscience, as conceived by Wolff, is nothing more than taking the moral point of view. Thus, whatever follows from the belief that the authority of one's state is morally legitimate, it cannot follow that the citizens of that state should regard themselves as having a duty to obey the laws of the realm even on the occasions when, all things considered, they oppose the law in conscience. Given what we mean by the moral point of view, a person cannot intelligibly regard himself as being required by morality not to do what he thinks is required of him by morality. We have then a curious state of affairs. The key philosophical question concerning political authority, as conceived by Wolff, admits of two interpretations. On one, the question is sensible, but surely not the one he had in mind. On the other, the question reflects his concerns all right, but on closer inspection, turns out to be incoherent.

II

As mentioned at the outset, Wolff is only one among many who adopt precisely the same confused starting point. Attention was focused upon his views because they are of considerable interest, not because they should be regarded as a particularly good specimen of intellectual pathology. Indeed, the fact that a thinker of such merit adopts an incoherent intellectual framework when analyzing the notion of morally legitimate political authority makes the issues to be dealt with in this section all the more pressing; namely, what are the questions we should be asking about morally legitimate political authority?; and what is it that has led so many philosophers to ask the wrong questions?

Frequently, when we ascribe authority to someone we are affirming his right to give orders to someone else and, correlatively, that other person's duty to obey those orders. For lack of a better phrase, authority in this sense will hereafter be referred to as *command authority*. Military officers, judges, football coaches, symphony orchestra conductors, and movie directors are just a few of the kinds of people who typically have command authority. This sort of authority, by its nature, arises in connection with social practices which assign roles to people such that in some roles the giving of orders constitutes appropriate behavior while in others, obedience to those commands is appropriate. *A*'s having command authority over *B* then is simply a matter of *A*'s and *B*'s occupying certain roles as defined by the rules governing a social practice in which they are both participants. Nothing more than this is entailed by the claim that *A* has command authority over *B*.

Consequently the rights and duties involved in command authority are not moral in nature. To establish the basis of *A*'s rights and *B*'s duties in virtue of *A*'s having command authority over *B* is to do nothing more than locate their respective roles within a given social practice. Independent argument is needed to legitimate this state of affairs from a moral standpoint. To take an obvious example of this, in an unjust war many soldiers find that, in conscience, they cannot justify obedience to certain commands which, looked at simply from the standpoint of their position within the military, it is quite clearly their duty to obey. In this sort of situation what we have is not a conflict of *moral* duties but rather, a conflict between two kinds of duties; moral and nonmoral.[7] Thus, one cannot answer the question "What is the moral basis of the state's authority?" in the manner appropriate to questions about the basis of someone or other's command authority. Morally legitimate political authority is not a species of command authority.

What is it then? What other type of authority is there besides command authority? On many occasions when authority is ascribed to someone it is not being asserted that he has a right to issue commands to others but rather, that, for some reason or other, he is worthy of respect. This kind of authority is referred to, for example, when it is said of someone that he is an authority or that he speaks with authority on a particular subject. To say these sorts of things about a person is not to suggest that he has a right to issue commands to others and that others have a duty to do what he says. Instead, it is to assert that his words on the subject in which he has expertise should be sought out and listened to carefully. This is the characteristic way in which one shows respect for someone who is an authority on a subject. It will be convenient to label authority of this sort *respect authority.*

Morally legitimate authority is a form of respect authority, for to say of a state that its authority is morally legitimate is to say that it is worthy of respect. More specifically, it is to say that the state does a tolerably good job of performing certain tasks which all rational men agree should be done but which they agree cannot be done on an individual basis. The Preamble to the United States Constitution, which was cited earlier, contains a good basic statement of these tasks; that is, the promoting of justice, domestic tranquility, the common defense, the general welfare, and liberty. There may well, however, be others. Some have argued that an important criterion for evaluating states is the extent to which they promote the fine arts and literature. Others have insisted that the state which takes no steps to preserve or establish a sense of community among its members is to be strongly criticized. Nonetheless, it is the satisfactory performance of these sorts of tasks which makes a state worthy of respect; and to say of a state that it is worthy of respect from a moral standpoint is to say that morality requires one to support it; that is, not to do anything that would contribute to its demise.

The key questions concerning morally legitimate political authority then are questions about the circumstances under which one is required by morality to support the state. Basically, they are the questions of exactly what tasks the state must perform and what standards of success it must meet with regard to each of these tasks to merit the support of the citizen. These are difficult matters. As mentioned earlier, there is by no means general agreement about what should be included in a list of the basic tasks of the state. Also, the question of how those tasks about which there is agreement should be ranked is far from settled. When we get into the matter of standards of success, things are even more problematic. Are there standards which all states must meet or do the standards vary depending upon factors such as level of technology, wealth in natural resources, and so forth? Or again, are there universal standards with regard to some of the tasks but not others? It is important to get as clear as possible on all these matters for they are the decisive factors in evaluating the justifiability of particular political revolutions or acts of resistance. Our views about whether a given revolutionary or resistance movement can morally justify its existence are a function of our moral assessment of the state it seeks to undermine. And this assessment, in turn, is in large part a function of our general views about what a state should do and how well it should do it.[8]

Conclusions about the basis of the state's authority then are primarily relevant for dealing with questions about the justification for revolution. They are only marginally relevant to questions about the nature and limits of one's duty to obey the law. If a person decides that the authority of his state is morally legitimate he thereby decided that he has a duty not to do anything that would contribute to its overthrow. But this does not, in itself, settle for him most of the difficult questions he might have about his duties with respect to law obedience. The question, "Why should I obey the law?" must be understood as a request for answers to any of a number of more pointed questions; for example, "Would Socrates have been guilty of any sort of moral transgression had he taken the opportunity to escape?" or "Is it ever justifiable in a viable constitutional democracy to register one's objection to a law by openly violating it (without the specific intention of testing its constitutionality)?" or "Is it really wrong to disregard some legal restrictions upon one's freedom when it is obvious to any rational person that it serves no purpose whatsoever?" Simply establishing that the authority of one's state is morally legitimate does not go very far in answering these sorts of questions. That one would be tempted to think that it does results from a faulty perception of the issues.

---

[7] On the distinction between moral and nonmoral obligation see Rolf Sartorius "Utilitarianism and Obligation," *The Journal of Philosophy*, vol. 66 (1969) pp. 67–81.

[8] All of these questions are far beyond the scope of this paper. I hope, however, to pay my respects to them in the not too distant future.

The distinction between command and respect authority is not only helpful in focusing upon the real issues of philosophical importance concerning legitimate political authority but also is useful in explaining what might account for the failure of so many philosophers to do so. It may be that philosophers such as Wolff have fallen into confusion over the concept of morally legitimate political authority as a result of committing two errors which are both rather easy to make. First, it may be that they have tended to think of morally legitimate political authority as being a form of command authority. Given the almost irresistable temptation to speak of the state itself as "the ruler" or "the soveriegn"—rather than reserving these phrases for reference to the *individuals* who wield power in the name of the state—it is no wonder that this mistake has been made. For if we picture the relations between those institutions which we call the state and those individuals whom we call citizens as analogous to the relations between a person who issues commands and those to whom he issues them, it is almost inevitable that questions about the basis of the state's authority will be taken as questions about its *right to command.*

Secondly, it may be that these same philosophers have failed to recognize that the rights and duties that arise in virtue of one party's having command authority over another are nonmoral in nature. This too is an error into which one might easily fall. It is, after all, a species of the more general error which has been made in other areas, of failing to distinguish between the justification of an act falling under a practice and the justification of the practice itself. A person who thinks that duties arising out of command authority situations are moral in nature believes that at times one can morally justify his actions by (a) showing that they were commanded by some person X and (b) showing that according to the rules governing the social practice in which he and X are engaged, it is appropriate for him to obey those commands. But this is, of course, mistaken. Showing (a) and (b) above would, at most, justify one's actions only in the eyes of other participants in the practice; that is, in the eyes of people who take the rules governing the practice as unquestioned standards of evaluation. It is thus not justification in terms of morality. To provide that sort of justification one would have to, at least, establish the moral legitimacy of the practice. This is why to believe that rights and duties arising out of command authority situations are moral in nature is to confuse justifying an action falling under a practice with justifying the practice itself.[9]

How do these two errors lead to confusion over the concept of morally legitimate political authority? If one thinks that such authority is a form of command authority, as would one who made the two mistakes described above, he would think that its basis is to be found in some relationship between state and citizen similar to those which typically obtain in command authority situations. But, as we have seen, the rights and duties created in these sorts of situations can conflict with the dictates of one's conscience. And so, it now seems that the rights and duties entailed by the morally legitimate authority of the state can also conflict with the dictates of one's conscience. That is, it now seems that there is a *prima facie* conflict between the idea that there is morally legitimate political authority and the principle that one ought to follow the dictates of his conscience. Also, since the right to command and the duty to obey are correlative, it seems that the above conflict is brought out when we ask the question, "Why is it that the citizen has a duty to obey the law even when he opposes it in conscience?"

---

[9] The major paper on this distinction is John Rawls's, "Two Concepts of Rules," *The Philosophical Review*, vol. 64 (1955), pp. 3–32.

In all candor it cannot be said that in every case where we find confusion over the concept of morally legitimate political authority there is hard core textual evidence of the two mistakes mentioned here. Nonetheless, they are mistakes which even a sophisticated thinker could easily make and they lead precisely to that confusion which plagues Wolff, and all the others who follow him in his formulation of the philosophical issues concerning political authority. Thus we have, at least, a plausible explanation of what might have led a number of highly competent thinkers into difficulties on this matter. In any event the point to be stressed here is that getting clear on the notion of morally legitimate political authority involves getting clear on such concepts as justice, freedom, and welfare; that is, on the goods which all rational men would agree it is the task of the state to deliver. It is not a matter of reconciling deeply held but incompatible beliefs.[10]

Illinois Institute of Technology

Received November 11, 1971

---

[10] The final draft of this paper was written during a one-semester period of partial released time granted me by the Department of Humanities and the Graduate School of the Illinois Institute of Technology.

The main thesis of this paper was discussed at length with Gary Thrane and Hugh Neuberger. An earlier draft of it was read at the work-in-progress seminar of the University of Illinois, Chicago Circle Philosophy Department. Thanks are due to all of the above for helping me to formulate my thoughts more clearly.

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 1972

VIII. WHY IS IT DIFFICULT TO DESERVE PRAISE?

EDWARD QUEST

"ALL great undertakings are risky and as they say, what is worth-while is always difficult."[1] What they say is strictly incorrect—tying one's shoes is at times both worth-while and easy to do—yet what they say points to a common sentiment which I believe to be correct. Namely, that it is hard to become an admirable, praiseworthy person; to lead a life which can be correctly described as non-hereditarily noble. The last line of Spinoza's *Ethics*[2] also points to this sentiment, as does the title of the evangelistic spiritual, "Without a Cross, There is No Crown." The explanation I offer of why this common assumption that it is difficult to lead a praiseworthy life is true can be given in a rather simple argument. First, if one has reasons for praising someone then one has reasons for encouraging that person or other persons to act in a similar way in the future. We may not wish to universalize the excellence we applaud, but we at least wish to have it reproduced. Secondly, there is no need to encourage someone to do what he already does easily. Hence, unless there is some difficulty in acting in a certain way, there is no reason to praise someone for acting in that way. Or in other words, if someone is deserving of praise, then someone finds it difficult to act in that praiseworthy manner.

My procedure for defending this explanation will involve first, examining an alternative analysis of why it is difficult to deserve praise, one offered by Aristotle. In the second section, I shall discuss the link between praise and encouragement and the link between encouragement and difficult behavior.

I

Aristotle answers the question, "Why is it difficult to do what is praiseworthy?" by comparing doing what is right[3] with successfully hitting a target.

> Now virtue is concerned with emotions and actions; and in emotions and actions excess and deficiency miss the mark, whereas the median is praised and constitutes success. But both praise and success are signs of virtue or excellence. Consequently, virtue is a mean in the sense that it aims at the median. This is corroborated by the fact that there are many ways of going wrong, but only one way which is right—for evil belongs to the indeterminate, as the Pythagoreans imagined, but good to the determinate. This, by the way, is also the reason why the one is easy and the other hard: it is easy to miss the target but hard to hit it.[4]

The positions I wish to understand and discuss are: (1) that good is determinate in a way in which its opposite is not; (2) that there are many ways of going wrong, only one way for doing what is correct, good, or right; and (3) this implies that it is difficult to do what is right.

The Pythagorean distinction mentioned in (1) between the determinate and indeterminate, between the [Form of the] Limited and the [Form of the] Unlimited,[5] is unfortunately for me an unclear distinction; so I shall attempt to understand what Aristotle is trying to summarize in (1) by discussing the target metaphor.

The first feature about the target which Aristotle seems to have in mind is that the bull's-eye is so small that there are an indeterminate number of ways of missing it; only one way to hit it. Consequently, it would be possible to know clearly that some ways of shooting would fail without knowing exactly how to succeed. Transferring this to its moral analogue, the consequence is in accordance with the well-known Aristotelian advice to face difficult moral problems (where one is unclear about what to do) by weeding away the alternatives that one clearly knows to be undesirable.[6] Just as one can know that a word is misspelled without knowing what is the correct spelling, so one can know, for example, that to seek revenge is wrong without knowing what exactly is the right thing to do. A second consequence of the target or bull's-eye being small is that it allows one to describe hitting it in more detail than missing it, i.e. "Joe hit the target" is a more graphic description than "Joe missed the target." A moral analogue of this is that the conditions, for example, sufficient for Joe being courageous are definite in a way in which the conditions consistent with Joe failing to be courageous are not.

---

[1] Plato, *Republic* 497d; trans. H. D. P. Lee (London, 1955), p. 260.

[2] "But all noble things are as difficult as they are rare." Spinoza, *Ethics* (New York, 1949), p. 280.

[3] For the reason that "It is difficult to do what is right" seems counter-intuitive if left unqualified, in the discussion that follows, the qualification will be that "right" and "good" are used in contexts of praise.

[4] Aristotle, *Nichomachean Ethics* 1106b, 24–32; trans. Martin Ostwald (New York, 1962), p. 43.

[5] Cf. Plato, *Philebus* 16c.

The thesis (2), and what I take to be its corollary, that there is an exactitude in describing moral success which is absent in describing moral failure, deserve scrutiny. But what is crucial, for my purposes, is that (2) does not explain why it is difficult to deserve praise. In spelling a word, e.g., "tennis," or in solving for x in simple algebraic equations, the number of inaccurate ways to choose does not make it difficult to choose correctly, even though the ways to fail are indefinite and the way to succeed is unique. If an archer is close enough, the bull's-eye is not difficult to hit. By analogy, I may do the right thing and it may be easy to do, e.g., decide to change my son's diaper, even though I am capable of hopping, skipping, or jumping out the back door. Thus, even if Aristotle is correct in holding that there is only one way to become a virtuous person, or to do that which is praiseworthy, he has not shown that this feature makes it difficult to become a praiseworthy person.

II

My explanatory inference may be restated as follows: (1) if A is deserving of praise for having ϕ-ed, then A or others are worthy of being encouraged to ϕ; (2) if A is worthy of being encouraged to ϕ, then it is reasonable to believe that A finds it difficult to ϕ; consequently, if A is deserving of praise for having ϕ-ed, then it is reasonable to believe that A or others find it difficult to ϕ. I shall approach the first conditional by considering a stronger claim made by P. H. Nowell-Smith in his *Ethics* that praise implies encouragement. "Praising is logically tied to approval; for if we heard a man praise something we could not wonder whether he approved of it or not unless we suspected him of being disingenuous or ironical; and it is logically tied in the same way to encouraging."[7] I believe that some of this is incorrect; and that one feature of praise should be emphasized, namely that praise implying encouragement is a stronger contextual inference than praise implying approval.

It is worth noting, first, that the possible suspicion Nowell-Smith refers to is not the same in the case of disingenuousness as in the case of irony. If Aunt Peg is insincere when she loudly applauds after a concert then she has praised without approving. I might suspect her of insincerity but I would not in this case suspect her of not praising. But whereas insincere praise is praise, ironical praise is not. If Aunt Peg sarcastically describes me as a terrific father then she is not praising me. Consequently, if I suspect her of being ironical then I am suspicious of whether she has praised and thus whether she has approved, not sure that she has praised and suspicious of whether she has approved. A similar point about irony contrasted with disingenuousness could be made with respect to the logical tie between praise and encouragement.

The converse of praise implying approval clearly does not hold; to approve of something is not necessarily to praise it. For one can hide approval while praise is a performance and therefore must be open in principle to inspection. The logical tie between praise and approval is rather like the relationship between praiseworthy and desirable. A praiseworthy action is desirable although what is desirable is not necessarily praiseworthy. It is desirable for me to take off my shoes each night but there is no point in praising me for doing this. Furthermore, the most desirable thing to do, i.e., the best thing to do, is not always praiseworthy. I step on a garden rake; the handle pops up and smacks my head; I then stagger to my bed and lie down while my wife calls the doctor. He states that what I am doing is the best thing I could be doing, which is not to suggest that I am deserving of praise.

The reason for stating that the move from praise to encouragement is stronger than the move from praise to approval is that encouragement contextually implies approval; whereas approval does not imply encouragement. If I encourage A to ϕ, then I approve of A's ϕ-ing. If I try to get Aunt

---

[6] Cf. Aristotle, *Nicomachean Ethics*, Bk. II, ch. 9.

[7] P. H. Nowell-Smith, *Ethics* (London, 1954), p. 169.

Peg to play the piano, then I approve of her playing. But if I approve of Aunt Peg wearing shoes to the market, I do not thereby feel that I should encourage her to do this.

Nowell-Smith claims that praise implies encouragement which I take to mean that if I sincerely praise *A* for having ϕ-ed, then (under penalty of being logically odd) I am encouraging *A* to ϕ. This is too strong. When one praises a feeble actor at his last curtain call, one may sensibly wish that he should never act again. Since the praise of a feeble actor may take place in private, the praise may not encourage anyone to act like this man. There are cases, to be sure, where praise does directly encourage whoever is being praised, e.g., yelling "encore! encore!"; but normally, I should think, praise expresses a sort of gratitude that indirectly encourages copies of the sort of activity that is being praised. For example, my applause at a concert expresses my enjoyment. With the implicit assumption that performers wish to please, such applause tends to encourage similar performances in the future. However, returning to the feeble actor who is praised upon his retirement without the praise being noticed by others, praise may take place quite easily without even indirect encouragement taking place. The more modest implication which strikes me as correct is that if someone is worthy of praise for having ϕ-ed, then someone is worthy of being encouraged to ϕ.

Let us suppose that it is reasonable to encourage someone, *A*, to ϕ. The following seems to be implied. First, ϕ-ing is something which it is reasonable to believe *A* can do intentionally. It is not reasonable, for instance, to encourage someone to meet a friend by accident. Secondly, it is reasonable to believe that ϕ-ing is something *A* should try to do. It is unreasonable to encourage someone to do what it is unreasonable for him to try to do. Thirdly, it is reasonable to believe that *A* will find or does find ϕ-ing difficult to do, and that this difficulty stems from *A* not being aware of the value of ϕ-ing. This last consequence arises from the fact that to encourage is not merely to urge, but to urge by stressing the value of the ϕ-ing which someone is being encouraged to do. An imperative directs but does not generally encourage. Encouragement may direct but its principal goal is to give someone reasons with which to overcome his or her reluctance or disinclination to do that which is thought, by those who encourage, to be valuable. The cheers of encouragement before the game supposedly help the team to "get up for the game" by making evident how high the stakes are, how important it is to win. If I encourage my son to study science, the encouragement is reasonable only if it is reasonable for me to believe that changing my son's beliefs about himself or about science will help him to overcome what is holding him back from studying science.

A counterexample for my first conditional will have to be a case in which *A* is deserving of praise for having ϕ-ed but one in which it is not reasonable to encourage *A* or others to ϕ. This might occur if *A* were deserving of praise for doing something which (1) could not be done again, or (2) if to do it again were not desirable, or (3) if one of the conditions that is implied by reasonable encouragement (three of which I mentioned above) does not hold.

Let us focus on the first two possible procedures for constructing a counterexample. Is anyone deserving of praise for being someone or doing something which will be eternally unique? Could anyone, for example, deserve praise for being the best nineteenth century philosopher? I think not in that the eternal uniqueness involved in being the best philosopher, for example, is not that which makes that philosopher worthy of praise; rather if he is worthy of praise, it is because of his excellence in being a philosopher—an excellence which can be duplicated. Considerations along this line suggest that not only is no one deserving of praise for becoming eternally unique but also one is deserving of praise only rarely because of the uniqueness of what one is doing, and then, only because that uniqueness is part of an excellence which can be reproduced.

In a chapter on the "vocation of the Hebrews" in which the discussion centers around what is meant by the claim that the Hebrews were chosen before other nations, Spinoza states:

> Every man's true happiness and blessedness consist solely in the enjoyment of what is good, not in the pride that he alone is enjoying it, to the exclusion of others. He who thinks himself the more blessed because he is enjoying benefits which others are not, or because he is more blessed or more fortunate than his fellows, is ignorant of true happiness and blessedness, and the joy which he feels is either childish or envious and malicious. For instance, a man's true happiness consists only in wisdom, and the knowledge of the truth, not at all in the fact that he is wiser than others, or that others lack such knowledge: such considerations do not increase his wisdom or true happiness.[8]

---

[8] Spinoza, *Theological-Political Treatise* (New York, 1951), ch. III, p. 43.

This brings out the reasonable thesis that being or becoming the best is not necessarily ideal, since the situation might be improved if others became as good as the one who is at present the best. In those cases where we do sensibly praise someone for being excellent in a unique manner, e.g., praising Cantor for being the first person to understand transfinite cardinals, it is only because being the first man to do this is an indication of being a brilliant mathematician. Similarly, we praise artists for their originality, but this is because their originality is an indication that they are imaginative, not that our reasons for praising them are novel. In short, originality is occasionally part of being excellent in some area, but being excellent in the area—mathematics, philosophy, art, etc.—is something that can be duplicated, and it is this excellence which is the fundamental reason why someone in such an area is deserving of praise. Consequently, I feel that it is impossible to construe a case where someone is deserving of praise for ϕ-ing (where ϕ-ing covers the reasons why the person is deserving of praise) but where no one can be encouraged to ϕ (for the reason that no one can possibly ϕ again or that to ϕ again would not be desirable.)

Let us now turn to the third way to attempt to construct a counterexample, i.e., by finding someone who is deserving of praise for having ϕ-ed, but where ϕ-ing is not something that (1) one can do intentionally, or (2) that one should do,[9] or (3) that is difficult to do. It is important for my purposes to show that one cannot be deserving of praise for unintentional conduct since unintentional conduct cannot sensibly be characterized as either difficult or easy. Suppose Joe is given, to his complete surprise, a great deal of money. Joe is not deserving of praise for having become a rich man. Bestowing a gift upon *A* does not make *A* deserving of praise. If we allow "gift" to be identified with "an unearned good thing," then I believe we can say that *A*'s "gifts" will not make *A* deserving of praise. One consequence of this is that since whatever *A* has at the beginning of his life cannot be earned by *A*, *A* cannot be deserving of praise for having a high I.Q., beautiful eyes, etc. More interestingly, if we assume along with classical theologians that God by his nature is unable to sin, and along with Pelagius that, "There would be no virtue in him who perseveres in the good if he had not had the possibility of going over to evil."[10] He will not be deserving of moral praise for being unable to sin. His inability is, by my terminology, a "gift." The general point here is that one can deserve praise for having ϕ-ed only if one could have been responsible for having ϕ-ed.

The general point above is stated weakly with the subjunctive in order to allow for cases of worth-while praise in which no one is responsible for having done what is worth-while. I am training my dog to sit; I force him to sit, or he accidentally sits, and I say "good! good! good!" in order to reinforce the desirable conduct. In teaching Joe to play the trumpet, I watch him struggle with a note until he accidentally succeeds in hitting it; at this point I praise him. Such pedagogic praise, where one praises good publicly observable conduct, even if it occurs accidentally, to encourage repetition and point out correct activity, is by no means improper or uncommon. Such praise is, however, disingenuous in that, returning to the example above, I am pretending to praise Joe as if he were responsible for playing the right note. Generalizing, when anyone deserves such pedagogic praise for having ϕ-ed unintentionally, it is only because one could ϕ and should ϕ intentionally.

Finally, is it possible to consider *A* worthy of praise for having ϕ-ed when ϕ-ing is not difficult for anyone? Again, my answer is, "no!," although it seems that even if the correct answer is "yes, rarely" this would be sufficient for explaining the soundness of the common sentiment that it is difficult to deserve praise. But hedging aside, the answer is "no." We praise a pianist for playing Brahms effortlessly, but this is because it is difficult to play Brahms effortlessly, i.e., it is difficult to become a pianist who plays Brahms effortlessly. We praise men for having the trait of being constantly considerate, not because considerate people must take pains to be considerate, but because it takes pains to become a considerate person. Remove the difficulty involved in ϕ-ing, and you thereby remove a reason for praising someone for ϕ-ing.

As an account of the origin of admiration, Adam Smith writes:

> When the sentiments of our companion coincide with our own in things of this kind, which are obvious and easy, and in which, perhaps, we never found a single

---

[9] The impossibility of *A* being worthy of praise for doing something which *A* should not have done seems obvious, so I shall treat only the first and third cases.

[10] Quoted in Walter Nigg's *The Heretics* (New York 1962), p. 134.

> person who differed from us, though we, no doubt, must approve of them, yet he seems to deserve no praise or admiration on account of them. But when they not only coincide with our own, but lead and direct our own; when in forming them he appears to have attended to many things which we had overlooked, and to have adjusted them to all the various circumstances of their objects; we not only approve of them, but wonder and are surprised at their uncommon and unexpected acuteness and comprehensiveness, and he appears to deserve a very high degree of admiration and applause. For approbation heightened by wonder and surprise, constitutes the sentiment which is properly called admiration, and of which applause is the natural expression.[11]

Smith correctly suggests in this passage that applause (or to use my term "praise") is approbation heightened by what is unusual. Approbation merely accompanied by what is unusual will not heighten the approbation—imagine the performance of a competent purple pianist; rather the unusual feature must be one which affects the degree of approval deserved. Such a feature is, I am insisting, the unusual degree of difficulty involved in being able to perform a certain sort of desirable conduct. Altruism, for example, is associated with praiseworthiness not so much for the reason that its consequences are unusually beneficial as for the reason that its practice is unusually difficult.

Consider the following exhortation from the Sermon on the Mount.

> If you love those who love you, what credit is that to you? For even sinners love those who love them. And if you do good to those who do good to you, what credit is that to you? Even sinners lend to sinners, to receive as much again. But love your enemies, and do good, and lend, expecting nothing in return: and your reward will be great . . .[12]

I have no inclination to think that the implicit reasoning in the passage above is the following, but it might be the following. Namely, it is easy to like those who like you and lend to those from whom you expect to borrow. Evidence for judging that it is easy lies in the fact that all men will act in this self-interested fashion, i.e., even a sinner. For the reason that it is easy, one does not deserve praise for such conduct, even though such conduct is correct. What is difficult as well as correct is to extend one's concern beyond one's personal interests, and this latter sort of conduct is deserving of praise. Because there is good inductive support for the generalization that that which everyone is inclined to do is not difficult to do, the reasoning above is sensible.

At this point, the sentiment underlying the paper may be restated: it is difficult to deserve praise. The underlying reason why one does not properly praise someone for having ϕ-ed when ϕ-ing is done with ease is that praise is conceptually tied to encouragement as outlined above. It is not that one usually praises with the intention of encouraging but rather that praise is reserved for what is uncommonly good, and what is uncommonly good does not commonly take place. Hence, men need to be encouraged to do that which will make them worthy of praise. The reason why it is difficult to deserve praise is that praise is reserved for those who do what is difficult.

California State College at Long Beach

Received August 5, 1971

---

[11] Adam Smith, *The Theory of Moral Sentiments* (Edinburgh, 1790), Part one, section one, ch. 4, p. 15.

[12] Luke 6, 32–35; R.S.V.

American Philosophical Quarterly
Volume 9, Number 4, October 1972

IX. "EXISTENCE" IN "THE EXISTENCE OF GOD"

ALASTAIR McKINNON

WHILE philosophers in the past have certainly disagreed about the existence of God, both supporters and opponents have tended to assume that this was in fact the primary question. In this brief comment I wish to argue that this question is at best secondary and that the fundamental one is whether we can give any sense to the word "existence" in the phrase "the existence of God." I assume that, if not, we should abandon the phrase and indeed the word as used in this context.

I can perhaps make my point with reference to Aquinas to whom my relation is at least somewhat ambiguous. Aquinas clearly saw that there were serious problems in ascribing even such qualities as goodness to God and he solved these by distinguishing between the *perfecio significata* and the *modus significandi*[1] or, as Copleston puts it, between the perfection signified by the term and by its mode of signification.[2] In certain passages, however, he appears to assume that no such distinction is required in the case of existence. Now it is true that he appears to hold that the ontological argument would work for God (if one may put the matter thus) and fails for us only because we are incapable of conceiving the divine essence. It is also true that Aquinas is much less prone to talk about "the existence of God" than many of his commentators. But these facts notwithstanding, he appears to assume that we can know, assert, and know what we mean when we assert the existence of God. Against this I want to say that "existence" has all the problems of "goodness" and some of its own as well.

Of course I realize that all the reasons are not on one side, and I should be the first to confess that there are good or at least obvious grounds for seeing the existence of God as an important if, on my own account, not primary item of religious belief. It seems, for example, obvious that if God does not exist, he could not in any way be relevant to anything. Similarly, it seems clear that if I do not believe in his existence I cannot think of him as relevant. Existence looks like the primary condition, the *sine qua non*, of everything else the believer wants to say. Without it, God could not have created or, rather, continue to create the world. He could not be involved in its destiny. He could not be a possible subject of worship or petition. If he had everything else but lacked existence, he would in fact lack everything. Existence seems to be that kind of thing. Hence the assertion "God exists" appears to be the primary element in religious belief. But I am not persuaded nor do I believe I should be. In fact, I think I can see stronger reasons on the other side.

My misgivings about the alleged primacy of this claim may emerge by casting this phrase in its interrogative form. Put as a question it asks whether God does in fact come up to or meet some independently defined and autonomous standard. It assumes that we know what we mean by "existence" even in this particular case and that we now wish to know whether God does in fact meet this test. It is like the question "Is Jones first class?" or, perhaps, "Does Jones pass?" Now it makes sense to ask if Jones is first class and it even makes sense to ask if he passes; at least there was a time when it used to. But, I submit, it makes no sense to ask if God passes, if he comes up to some standard which we have quite independently derived and formulated. And, I further submit, anyone with a clear sense of the way "God" has been used in religious belief will, at least on reflection, recognize this fact. He will recognize that we cannot say that God measures up to some independent and autonomous standard which we have set for him. Of course, someone might intend the word "existence" in a radically new and different sense somehow derived from insight into the nature of God. I do not deny this possibility but observe that "God exists" would then be a tautology and have a structure and force quite different from that ascribed to it by the tradition.

Let me put this reservation, for this is perhaps the right word, in another way. The word "existence" might be used here in either some new or in the old, familiar, rooted-in-the-mundane-world sense. If the former, I suggest we abandon the word together with its historic circumlocutions. If the latter, I suggest that neither the assertion nor the mood in which it is made is good enough. It may well be true that God is not good in the sense in which a man may be good but it is yet more evidently true that he does not exist in anything like the sense in which a stone exists and those who see no difficulty with the traditional claim can scarcely be thinking of anything worthy of the name God.

---

[1] E.g., Aquinas, *Summa Theologica*, I, 13, 3.

[2] F. C. Copleston, *Aquinas* (Baltimore, 1955), pp. 130f.

Having confessed these misgivings, let me now attempt to state my case with specific reference to the word "existence" and its familiar variants.[3] Hopefully this will bring us closer to something approximating a solution.

There are, I suggest, three possible types of entities of which one might be tempted to say that they exist:

(1) the ordinary and familiar objects of the spatio-temporal world.
(2) the stuff (e.g., the elements of some of the pre-Socratics) of which these ordinary objects are allegedly made.
(3) Being or Reality which is the source or ground of all that is.

It is important to be clear about the various motives we might have for saying of any or all of these entities that they exist. And it is equally important to be clear as to what we could possibly mean by such claims.

It is scarcely possible to make too much of the fact that we only very rarely use the word "exists" in connection with objects of the first class. Put another way, this word scarcely has a normal or common use context. But we can perhaps imagine situations in which we might be provoked into its use. For example, "My dog exists" ("I really do have a dog and the story is not a fabrication"), "A flag exists on the moon" ("No, it was a real flag and not a trick done with television cameras"). Even, I suppose, "The word 'existence' exists in the dictionary." Note, however, that the assertion is in reply to an essentially frivolous doubt and, more to the point, that it is voiced just because one can imagine the possibility of its falsehood. I say "My dog exists" because I understand that someone can intelligibly doubt it. In short, we would say of physical objects that they exist just because we are so fully aware that their existence is fortuitous and contingent. This is an essential feature of the so-called ordinary use of "exists" and is one which precludes even its analogical use in conjunction with at least the more interesting uses of "God." The point is that our motives for using "exists" in the first case are precisely opposite to those for using it in the second.

The second use recalls a very familiar move in the history of philosophy. Having emphasized the extent to which ordinary physical objects are subject to change and decay, certain philosophers have been prone to claim that the object or objects of their attention are subject to no such vissicitudes. And, they add, it is only of such elements that one might rightly say that they exist. The argument is familiar and, to a point, plausible. Physical objects do pass away and those with a proper jealousy for "exists" might be unwilling to see it squandered upon that which only happens for the moment to be. I believe that this objection has some merit and that, seriously pondered, it may help us to understand why this word is so rare; why, in fact, some languages appear able to dispense with it altogether. But I suggest that in the end it too collapses. For the test is not whether something may cease to be but rather whether we can consistently imagine it as so ceasing; in this case, whether we can imagine the world as having a different ultimate constitution than that which it is alleged to have. And plainly we can imagine that this might be the case. Hence while the so-called stuff of the world may be free of what we might term immediate contingency it is nevertheless plainly infected with an ultimate one. Hence uses of "exists" which are anchored in this second context are no better suited for conjunction with "God" than those which spring from the first.

The third possible object, Being or Reality, presents a quite different type of problem. In Tillichian language, Being is not *a* being but rather the ground of all possible beings. It is Being itself, that whose passing or ceasing to be we cannot either fancy or imagine. It is quite unlike those fictive and problematic entities which we encounter at the end of each of Aquinas' five ways; in fact, it is something many might be tempted to call God. But the very reasons which might make us want to do so should, I claim, also make us unwilling to say that he exists. Of course, if we were to speak of the term as opposed to its mode of signification, if for a brief moment we imagine that we are as God, we might perhaps want to say that Being or Reality does exist (i.e., we would perhaps be in a position to know what we meant in so saying.) But precisely because Being is not a being, because God is not even conceivably contingent, because we cannot think of him as not existing, we would, paradox of paradoxes, not want to say that he exists. This word is anchored in the world of contingency, immediate or ultimate, and cannot be freed to go easily in pairs with Being, Reality, or God. I cannot say "God exists" precisely because I cannot even think of him as not existing. I cannot do so for the same reason that I cannot say "A circle is circular" or "A triangle is triangular." I go back to a central utterance in Wittgenstein, a remark which, incidentally, occurs within the context of a discussion of the conditions under which one might attribute existence: "There is *one* thing of which one can say neither that it is one metre long, nor that it is not one metre long, and that is the standard metre in Paris."[4] But I go back only to go ahead: "There is *one* thing and only one thing (I did not say, 'in the world') of which it makes no sense to say that it exists, and that is (the) God who could not not exist."[5]

---

[3] I have formulated much of my argument with specific reference to the word "exists" but would claim that expressions such as "There is a God" are subject to the same objections.

I turn now from this cursory examination of real and possible subjects for "exists" to another area which anyone proposing to take ordinary language seriously must consider carefully. The first is the Bible and the second is the worship of the Church as typified, for example, in the Apostles' Creed.

It is constantly assumed and sometimes even alleged that the Bible asserts the existence of God. The best and indeed only evidence for this view is found in the Psalmist's celebrated utterance "The fool hath said in his heart there is no God."[6] Now it is plain from this that the writer assumes that only a fool would say that there is no God from which however it does not at all follow that a wise or good man must say that there is. Certainly, it does not at all follow that such a man must assert the existence of God. Indeed, much of the real force of this remark appears to lie in the phrase "in his heart" which seems in fact to point away from assertions altogether. But more about this in a moment.

It is perhaps worth remarking that the Apostles' Creed does not begin with an assertion of the existence of God but rather with something quite different, *viz.*, the professing of belief in God. Elsewhere I have tried to explicate some of the elements which are packed into and perhaps partly camouflaged by this formula.[7] For the moment I simply report that, in my view at least, it cannot be interpreted as equivalent to an assertion of the existence of God in the sense which I am here attempting to question.

It would be quite wrong to suppose that these two traditions are different or, indeed, opposed. In fact, the Apostles' Creed simply makes explicit an emphasis which characterizes biblical thought as a whole. That emphasis is on personal trust and commitment; it is on a way of life as distinct from any primarily intellectual affirmations which undergird this way or which may be thought of as doing so. In fact, it is on a trust which is very much like commitment to another person. It is then relevant to ask whether and in what sense such commitment involves belief in the existence of its object.

Now at first glance it seems that nothing could be plainer than that my trust or belief in John Brown presupposes another and earlier belief in his existence; in fact, I could hardly be said to trust him unless I already believed in his existence. It is surely this self-evident truism which has prompted so many to assume that the existence of God is the first or primary item in religious belief. But this concerns the existence of God only in a trivial sense and in fact the matter is much more complex than this simple account assumes. While I could hardly be said to trust someone in whose existence I did not believe, neither could I be said to believe in or trust someone (in any sense at all counterpart to the religious use) if, in the matter at issue, I already have certain knowledge that he will behave as I hope. If, for example, I know that he will not reveal my secret under the most extreme duress, it would be misleading simply to say that

---

[4] Wittgenstein, *Philosophical Investigations* (Oxford, 1953), §50.

[5] I have not discussed the alleged and perhaps possible use of "exists" in connection with mathematical entities because I believe such use to be non-ordinary and, more important, highly problematic. Further, it seems clear that the alleged non-contingency of such entities is not at all comparable to that of God.

[6] It is true that the words ὅτι ἔστιν occur in Hebrews 11:6b with reference to God. I would argue however that this involves a neutral and finally uninteresting sense of "existence" and that the use in question is certainly not grist for any of our modern philosophical mills.

[7] Alastair McKinnon, *Falsification and Belief* (New York, 1970), especially chaps. III and IV.

I believe or trust that he will not do so. The truth is that I know he will not and here as elsewhere we should say what we mean. Now we do sometimes know, if perhaps in some weaker sense, that God will do thus and so. I have attempted to explicate this element in religious belief and have described it as involving the assertional use of our claims and the determinate use of the relevant or key term. Such uses are important but it is not these with which we are here primarily concerned. The focus of religious belief is not the God I can conceive nor, in general, the one whose actions I can anticipate. The real focus of such belief is not the idols I form and must form but rather he who is beyond and yet the object of all my idols. To say "I trust John Brown" *may* be to presume the existence of someone roughly corresponding to my conception of John Brown; to say "I believe in God" is, in the interesting cases at least, to presume nothing of this kind. But this brings us to the two senses in which "God" might conceivably be used in the claim "God exists."

Someone making this claim might use "God" in what I shall call the determinate use. By this I mean that he can spell out what he means or, perhaps more accurately, can say what he now has primarily in mind in using this term. He might, for example, suppose that God is an indulgent father and intend something like "An all powerful, indulgent father exists." But everyone knows that this conception is wrong and inadequate. More importantly, just about everyone knows that *any* conception is almost equally wrong and inadequate. But to say that a conception is wrong is to say that its object does not exist. From this it follows that any use of "God exists" in which the determinate use of "God" is to the fore must necessarily be false. Further, any serious religious believer must recognize it as such.

Nor is the situation any better with the second use of "God" which I shall call heuristic. But here the difficulty is very different. While there necessarily is an object corresponding to this use, the believer cannot, on his own admission, specify its nature. Here he does not have a conception whose object he can claim to exist. Indeed, all he knows is that there must be some object corresponding to this use. Of this God he can say only that he could not not be. And it is precisely this which prevents him from saying that he exists.

To put the matter very briefly. Existence statements appear ultimately to be concept-correspondence claims; to say "*X* exists" seems to be to assert that there is somewhere an object corresponding to my conception of *X*. But on this analysis neither use of "God exists" will do. That involving the determinate use of "God" is necessarily false and that involving the heuristic use has as grammatical subject a term which is plainly not a concept in anything like the traditional sense.

I have attempted to state my case within the framework of traditional philosophical assumptions; to show that, granted these assumptions, one cannot properly speak of the existence of God. That I do not myself accept them is another matter; the fact remains that one must assume at least much of the language of one's community in order to communicate. But I doubt that I would wish to reinstate this traditional claim even were I to succeed in effecting a real revolution in my own philosophical outlook. Indeed I suspect that the use of "exists" in this connection is at the very heart of the tradition which has to be overthrown and, given a new set of assumptions, that one might produce yet stronger arguments for rejecting this claim than those which I have suggested in this paper.

McGill University

Received July 8, 1971

AMERICAN PHILOSOPHICAL QUARTERLY
Volume 9, Number 4, October 1972

X. THE PARADOX OF DETERMINISM

A. AARON SNYDER

UNDERLYING various forms of determinism is the idea that every event in the universe is so determined by its causal antecedents that, given the occurrence of the causal antecedents of an event, that event could not fail to occur. This doctrine of universal causal determinism has a certain sweeping grandeur which survives successive attempts at analyzing it into more specific views. It suggests a picture of a world like an enormous clockwork precisely regular in its operations, and hopefully, simple in the ultimate principles which govern those operations. It conveys a feeling that our world, no matter how complex it may appear to be, is ultimately intelligible, in principle at least. It holds out the hope that with ever increasing knowledge, we could come to explain and even predict without limit the occurrence of phenomena which we do not currently understand. The attraction of this idea is, consequently, that any deficiency in our understanding of the way our world works is not in principle irremediable; it is a function solely of our limited knowledge of things and not of the way things are in themselves.

But there is an intuitive repulsion in this idea, as well, for it may be used to generate some of the more distressing forms of the Free Will Problem. If the world is like an enormous clockwork, then we are like parts of the clock. If every event is determined by causal antecedents, then supposedly, so is the behavior of persons. And if — given the occurrence of the causal antecedents of our behavior — we could never have done other than we did, then a distinguishing feature of persons seems to evaporate; namely, it does not seem that persons can be moral agents.

Of course this doctrine of universal causal determinism is too general as it stands. It does not specify what is to count as an event, it does not specify what sorts of antecedents are to count as causal, and it only hints at the nature of causal connection. This doctrine demands specification, analysis, and piecemeal treatment, but even in this general form, it has considerable philosophic interest. I want to approach it in a spirit of corresponding generality and suggest that, even as underspecified as it is, it has self-refuting implications.

* * *

It is an old notion that trying to convince someone of the truth of causal determinism smacks of self-refutation. Perhaps because this idea is so old, or perhaps because it has only been hinted at, it is usually treated as a bit of quaintness, not to be considered seriously among the arguments that weigh in the Free Will Problem. It seems to me, however, that this notion is the product of perceptive logical olfaction and I would like to isolate its source. I will try to show that it is self-refuting to suppose that there *could be* a proof of the doctrine of universal causal determinism.

This supposition needs some explanation. First, by the doctrine of universal causal determinism (which I shall call simply "the Determinist Doctrine") I mean the general doctrine mentioned above; *viz.*, that every event in the universe is so determined by its causal antecedents that, given the occurrence of the causal antecedents of an event, that event could not fail to occur. This is often put more succinctly by saying that (1) every event is caused and (2) the cause of any event is *physically sufficient* for its occurrence. I shall employ the terminology of "physical sufficiency," but only as shorthand. Those who are justifiably suspicious of this terminology may replace it by the longer expression for which it is shorthand; *viz.*, take "x is physically sufficient for y" to mean that if x occurs, y cannot fail to occur. Although the notion of physical sufficiency is hardly very clear, at least this much is clear about it: it is weaker than logical sufficiency, but stronger than mere material sufficiency. That is, causes do not logically entail their effects; but to say that x caused y is to say more than merely "It is not the case that x occurs and y does not occur." That physical sufficiency is not identical with either logical sufficiency or material sufficiency can readily be seen by considering that it is a relation between event-like entities, while they are relations between propositions.

The word "event" in the statement of the Determinist Doctrine is to be understood as a surrogate for a great variety of things: occurrences, situations, states of affairs, properties — in short, almost any contingent feature of the universe which could be said to occur, obtain, or exist. Although the boundaries of the class of events under this generous construal is not well-defined, we have a fairly good idea of what some of the more attractive candidates for membership are.

What is important, in order that the Determinist Doctrine figure in generation of the "Free Will" problem, is that the class of events should specifically include things done as well as things which merely occur. That is, what it says about events is intended to extend equally to the behavior of persons.

By a "proof" I mean any argument or reason for believing a proposition *p* which shows *conclusively* that *p* is true. Hence, it will be possible to prove *p* — or provide conclusive argument or reason for *p* — only if *p* is true; though of course, *p* may be true even when we cannot prove that it is.

In saying that the supposition that there is a proof of the Determinist Doctrine is self-refuting, I do not mean that the doctrine is itself inconsistent. The inconsistency resides instead in the complex concept, "proof of the Determinist Doctrine;" that is, the Determinist Doctrine is inconsistent with there being proofs, conclusive arguments, or conclusive reasons for believing anything, and hence, trying to prove that doctrine is a self-refuting activity.

This inconsistency arises out of the fact that the universal operation of physically sufficient causes would leave no room for the conceptual sufficiency of reasons. If *R* is conclusive reason for believing *p*, then *R* would make a rational man (i.e., one who is in a perfectly rational state and is informed of all the relevant facts) believe that *p*; that is, one has not provided conclusive reason for believing *p*, unless it is the case that what he said would make such a rational man believe that *p*. But, as we understand the notion of conclusive reason, this "would" expresses a conceptual relation between the propositions *R* and *p*, not a physical relation between believing *R* and coming to believe that *p*.

Moreover, at least for a great many cases, I ought to believe *p* or do *x* only if I have good reason for believing *p* or doing *x*. But (as most determinists would urge in the course of generating the Free Will Problem) there is no clear sense in saying that one ought or ought not to believe or do what one is causally determined to believe or do. These and other conceptual relations among what one has reason for believing, what one would believe, and what one ought to believe are significant features of our concepts of reason and proof, but would be vacuous in a deterministic universe.

Presumably, one of the entailments of the Determinist Doctrine is that what one believes is, like every other event, state, etc., *causally* determined. If, in a deterministic universe, I believe that *p*, I do so entirely because of the causal antecedents of my coming to believe that *p*. Whether I come to believe that *p* would be independent of whether I have (what we should call) reason or proof for *p*; so, specifically, whether I come to believe the Determinist Doctrine would be independent of my having reason for it or proof of it. Thus, it would be senseless to attempt to convince someone of the truth of that doctrine by providing reason, argument, or proof in its support.

* * *

But what if my having reason for believing the Determinist Doctrine *is* what causes me to believe it? — Then it would not be senseless to try to convince someone of that doctrine by arguing for it. So one might defend the Determinist Doctrine against this paradox by claiming that reasons are causes, and hence, that my having reason for believing *p* is just what causes me to believe that *p*.

This is a plausible response. After all, we do often speak of doing things and believing things *because* we have certain reasons for doing *x* or for believing *p*. Moreover, we often answer such questions as "What made him do that?" (and perhaps, though this is less likely, "What caused him to do that?") by supplying his reasons for his belief or his behavior. It does seem that in such cases we are treating his reasons as the cause of his belief or behavior.

This provides little support for the Determinist Doctrine, however, since both "reason" and "cause" have undergone a shift in sense. First, *his* reason for believing *p* need not be *a* reason for believing *p*. And when we say that someone *has* reason for believing *p*, we are saying more than that there is something which is *his* reason for believing it; in addition, *his* reason must be *a* reason (*cp.* "He had what he *believed to be a* reason.") So from the fact that his reason for believing the Determinist Doctrine is the cause of his believing it, it does not follow that what caused his belief is *a* reason for believing the Determinist Doctrine.

If he has reason for believing the Determinist Doctrine, and is a rational man, his having that reason (which is also, now, *a* reason) may cause his belief. But this is not a sense of "cause" according to which causes need be physically sufficient. Consequently, it would be inconsistent with the Determinist Doctrine to count one's having reason for believing the Determinist Doctrine as the cause of one's believing it.

If having reasons for believing *p* were physically sufficient for believing *p*, then it could not be the case, as it is in fact, that we have reasons both for and against believing or doing the very same thing. Nor could it be the case, as it is in fact, that of two persons who have the very same reasons for believing *p* or for doing *x*, just one of them believes *p* or does *x*. Moreover, we speak of "accepting and rejecting reasons" and of "more or less compelling reasons." But if reasons were physically sufficient, all reasons would be equally compelling and none could be rejected.

One could insist that, even if having reasons is not always physically sufficient to produce belief, nonetheless, my having conclusive reasons for believing *p* or doing *x* is physically sufficient for my coming to believe *p* or my doing *x*.

But this claim is not true either. I might have the very best of reasons for believing something or for doing something, and yet as a matter of fact, not believe or not do it. Such behavior might indicate foolishness or weakness of will, but it does not indicate that I have a supernatural ability to resist physically sufficient causes.

The status of *R* as a reason for *p*, or even its status as a conclusive reason for *p*, does not require that someone's having *R* be *physically* sufficient to produce belief in *p*. Indeed, if we require that reasons have physical sufficiency, then nothing that has been given as reasons for believing the Determinist Doctrine would count as reasons for accepting it, since many persons who are aware of these alleged reasons have, nonetheless, not come to believe it.

Thus, the fact that we sometimes count reasons as causes — far from supporting the Determinist Doctrine — actually counts against that doctrine. If in the claim that reasons are causes, "reasons" is meant in an ordinary sense, then it follows that not all causes are physically sufficient, and the Determinist Doctrine is false. But if, on the other hand, this is meant in a sense that requires physical sufficiency, then nothing which has been given as reason for accepting the Determinist Doctrine can count as reason. So, if reasons are causes, the Determinist Doctrine is either false or unsupported.

If one deliberately grasps the second horn of this dilemma, by maintaining that the concept of "reasons" is one which really has no use in our world, then he must also recognize that it would be self-refuting to try to provide reasons, arguments, or proofs for determinism.

* * *

A determinist might agree that having reason, even conclusive reason, need not be physically sufficient to produce belief, and yet insist that having reason does play a causal role in belief. That is, he might maintain that having conclusive reason is *part of* a physically sufficient cause; the other part is that the person who has that reason be in a biophysical state of a sort which is ordinarily indicated by saying of the man that he is in a rational state and is informed of all the relevant facts. Such a rational man who has conclusive reason for believing *p*, could not fail to believe *p*. Thus, the fact that some persons who are aware of the reasons for believing the Determinist Doctrine have nonetheless not come to believe it can now be explained by saying that either they are not fully informed, or they are not rational men, or at least, in refusing to accept determinism, they are not acting rationally. By this device, one may hope, the Determinist Doctrine and the possibility of proving the Determinist Doctrine are rendered mutually consistent.

Unfortunately, this device also trivializes the Determinist Doctrine. It emerges that the physically sufficient condition for *A*'s believing that *p* is that *A* have conclusive reason *R* for believing *p* and also that *A* is a rational man. (*NB*. *A* must be in a perfectly rational state and must be informed of all relevant facts, lest the slightest deviation from perfect rationality or the slightest lack of relevant information frustrate the physical sufficiency of his having *R*.)

Given the logical relation which exists between "being a rational man who has conclusive reason for believing *p*" and "coming to believe *p*," this view is approximately as cogent as the view that being a female fox is physically sufficient for being a vixen or that being the successor of 2 is physically sufficient for being the square-root of 9. On this account, physical sufficiency has collapsed into logical entailment and the Determinist Doctrine is reduced to a fascinating tautology. Thus the

Determinist Doctrine can escape paradox only at the cost of total trivialization.

* * *

Though it is somewhat unsatisfying, one could still hold that even if it be logically impossible to prove the Determinist Doctrine, the proposition which it expresses could nonetheless be true.

My response to this would proceed along two lines. First, it is not usually claimed that the Determinist Doctrine expresses the sort of commonsensical position which does not stand in need of argumentative support. The usual view is that it is a position for which such strong arguments can be given that, unless we are "unreflectively committed to the metaphysical prejudices of our Stone Age ancestors," we must accept it on the strength of those arguments. I have tried to show that the supposition that anyone could provide such arguments, much less that anyone has, is a self-refuting supposition.

Moreover, though I am uncertain of this, I think that the only propositions which are *logically* incapable of proof are those which are false. In any case, this seems the most plausible explanation of why this particular proposition is unprovable.

The University of Wisconsin

Received April 14, 1971

BOOKS RECEIVED

Joseph Agassi, *Faraday as a Natural Philosopher* (The University of Chicago Press: Chicago, 1972), 359 pp. $12.50.

Elizabeth Anscombe and Peter Thomas Geach (eds.), *Descartes Philosophical Writings* (Bobbs-Merrill: New York, 1971), 303 pp.

Richard J. Bernstein, *Praxis and Action* (University of Pennsylvania Press: Philadelphia, 1971), 344 pp. $3.95.

Roger C. Buck and Robert S. Cohen (eds.), *PSA 1970 In Memory of Rudolf Carnap, Boston Studies in the Philosophy of Science*, Volume VIII (D. Reidel Publishing Co.: Dordrecht, 1971), 615 pp.

Warren D. Goldfarb (ed.), *Jacques Herbrand Logical Writings* (D. Reidel Publishing Co.: Dordrecht, 1971), 312 pp.

R. M. Hare, *Practical Inferences* (Macmillan: London, 1971), 120 pp.

Glenn Langford, *Human Action* (Anchor Books: New York, 1971), 137 pp. $1.45.

Hugues Leblanc and William A. Wisdom, *Deductive Logic* (Allyn and Bacon: Boston, 1972), 367 pp.

Alexandre Matheron, *Le Christ et Le Salut des Ignorants Chez Spinoza* (Aubier Montaigne: Paris, 1971), 284 pp.

Jitendra Nath Mohanty, *The Concept of Intentionality* (Warren H. Green, Inc.: Missouri, 1972), 213 pp.

Clarence Morris, *The Justification of the Law* (University of Pennsylvania Press: Philadelphia, 1971), 214 pp.

Henri Mougin, *La Sacra Famiglia dell'esistenzialismo* (pubblicazioni dell'universita di urbino: Italy, 1971), 274 pp.

Octav Onicescu, *Principles De Logique et de Philosophie Mathematique* (Editions de l'Académie de la Republique Socialiste de Roumanie: Bucharest, 1971), 229 pp.

Hans Regnell (ed.), *Readings in Analytical Philosophy* (Läromedelsförlagen: Sweden, 1971), 195 pp.

Richard Rudner and Israel Scheffler (eds.), *Logic & Art* (Bobbs-Merrill: New York, 1972), 330 pp. $10.00.

Frederick Sontag, *How Philosophy Shapes Theology: Problems in the Philosophy of Religion* (Harper & Row: New York, 1971), 495 pp.

Herbert Stachowiak, *Rationalisms im Ursprung* (Springer-Verlag: Wien and New York, 1971), 348 pp.

Paul W. Taylor (ed.), *Problems of Moral Philosophy* (Dickenson Publishing Co.: California, 1972), 559 pp.

Fritz-Joachim von Rintelen, *Contemporary German Philosophy* (H. Bouvier u. Co. Verlag: Bonn, 1970), 177 pp.

CORRIGENDA

Volume 8 (1971)

HOWARD KAINZ AND FREDERICK WEISS : "Recent Work on Hegel," (pp. 203–222).

Page 218, col. 2, line 3: delete footnote 98

line 4: change footnote 99 to 98

line 7: change footnote 100 to 99

line 7: add: and a translation of the *Phänomenologie des Geistes* has been commissioned.[99]

Volume 9 (1972)

SAM COVAL AND D. D. TODD: "Adjusters and Sense-Data," pp. 107-112.

Page 112, col. 1, line 11 from bottom should read:

> We *could*, that is, introduce what we now call hallucinated daggers into our conceptual scheme independently of the concept of dagger just as we *could* introduce ersatz coffee independently of coffee.

VOL. LXXXI No. 324 **MIND** October 1972

Edited by Professor DAVID HAMLYN

- I Logical Possibility: G. SEDDON
- II Definitions and 'Clusters': D. E. COOPER
- III How to Tell Your Friends from Machines: G. ROBINSON
- IV Religious Belief and Philosophical Analysis: D. A. McDOUGALL
- V Psychology and the Philosophy of Science: G. WESTLAND
- VI On Thinking: J. HARTNACK
- VII The Intelligibility of Wants: A. J. WATT
- VIII Discussions:
  - The Connection Argument: G. ISEMINGER
  - Whether the Theory of Family Resemblances Solves the Problem of Universals: J. W. THORP
  - Urmson on Evaluation from a Point of View: J. W. NICKEL
  - Vagueness and Colour Predicates: C. BOYNE
  - Austinian Ifs: R. J. FOGELIN
  - Reply to Ruth Anna Putnam: W. T. SCOTT
  - A Doubt about the Normative Theory of Belief: D. PUGMIRE
  - A Solution to an Ethical Paradox: F. J. LEAVITT
  - Description and Evaluation: D. CONWAY
  - Hedman on Explanation: J. E. WHITE
  - On a New Escape from Logical Determinism: T. CHAPMAN
  - A Note on Hampshire's Analogy: J. BELFORD
- IX Critical Notes:
  - The Bounds of Sense: an Essay on Kant's Critique of Pure Reason: by P. F. Strawson. W. CERF
- X New Books
- XI Notes.

Published for the MIND ASSOCIATION by BASIL BLACKWELL

Single Issues: 40p. Annual Subscription: £1.25 net

RATIO edited by STEPHAN KORNER

Volume XIV December 1972 No. 1

RODERICK M. CHISHOLM and ROBERT G. KEIM	A System of Epistemic Logic
K. E. TRANØY	'Ought' Implies 'Can': A Bridge from Fact to Norm
JAN SRZEDNICKI	The Transcendental Impossibility of Solipsism
R. S. WOOLHOUSE	From Conceivability to Possibility
WOLFGANG CARL and ROLF P. HORSTMANN	Knowing and Claiming
DONALD SIEVERT	Does Prichard's Moral Philosophy Rest on a Confusion?
KEITH GRAHAM	Ifs, Cans and Dispositions
REVIEWS:	
R. M. HARE	*The Object of Morality* (G. J. Warnock)
W. H. WALSH	*Early German Philosophy—Kant and his Predecessors* (Lewis White Beck)
EDGAR MORSCHER	*Rechtslogik. Versuch einer Anwendung moderner Logik auf das juristische Denken* (Ota Weinberger)

Price £1.75 net. Annual Subscription £3.00 net.

BASIL BLACKWELL
Oxford, England

CANADIAN JOURNAL OF PHILOSOPHY

PUBLISHED BY THE CANADIAN ASSOCIATION FOR PUBLISHING IN PHILOSOPHY

Editorial Board—John King-Farlow (Alberta), Kai Nielsen (Calgary),
Terence Penelhum (Calgary), William Rozeboom (Alberta) – **La Redaction**

VOLUME II NUMBER 2 DECEMBER 1972

COLOR QUALITIES AND REFERENCE TO THEM	N. L. Wilson
THE CAUSALITY OF GOD IN SPINOZA'S PHILOSOPHY	A. J. Watt
COMMENT ON WATT	Errol Harris
THE JUSTIFICATION OF LIBERALISM	D. A. Lloyd Thomas
PRIMARY AND SECONDARY QUALITIES	Keith Campbell
SOME REFLECTIONS ON QUANTIFIED EPISTEMIC LOGIC	Robert C. Coburn
A SHORT REFUTATION OF ETHICAL EGOISM	Richmond Campbell
THE LANGUAGE-GAME VIEW OF RELIGION AND RELIGIOUS CERTAINTY	James Kellenberger
FREEDOM AND PLASTIC CONTROL	Alastair Hannay

Annual Subscription – Institutions $12; Individuals $10; Students $5 – **Abonnement Annuel**

Subscriptions to – Roger A. Shiner (Managing Editor) – **pour Abonnements**

Manuscripts to – Charles G. Morgan (Executive Secretary) – **pour Manuscrits**

Department of Philosophy, University of Alberta. Edmonton, Alberta, T6G, 2E1, Canada

---

DIALOGUE

Canadian Philosophical Review—Revue Canadienne de Philosophie
Editors: Venant Cauchy and Martyn Estall

Vol. XI—1972—No. 3

ARTICLES

Two Dogmas of Linguistic Empiricism	JOHN KING-FARLOW
Philosopher, c'est apprendre à mourir	ADÈLE CHENÉ-WILLIAMS
On Refusing to Play the Sceptic's Game	KAI NIELSEN
La Commune de Paris, Fête populaire	HENRI LEFEBVRE

NOTES—DISCUSSIONS—REVIEWS

Subscriptions: $10.00 a year to individuals; $12.00 to Libraries
Payable to the Canadian Philosophical Association in
care of Norman J. Brown, Department of Philosophy,
Queen's University, Kingston, Ontario, Canada

METAPHILOSOPHY

edited by
Terrell Ward Bynum and William L. Reese

Vol. 3, No. 4 CONTENTS October 1972

Is empiricism an ideology? KAI NIELSEN
Philosophy—practice and theory: a venture into the sociology of philosophy. JANNA L. THOMPSON
The concept of speculative philosophy of history. ROLF GRUNER

Reviews

K. T. Fann. *Wittgenstein's conception of philosophy* WILLIAM G. LYCAN
The Philosopher as Teacher: Articles, Comments, Correspondence.
A new medium for teaching philosophy. DAVID WEST.
The black box as an aid in teaching philosophy. ROBERT L. CAUSEY

Notes

A correction to a review of Prior's *Objects of thought.* LEONARD LINSKY

Books Received

Review Editor: Harold Morick Managing Editor: James A. Thomas

Annual Subscription: $15.00 (*Institutions*) $10.00 (*Individuals*) *Single copies:* $5.00 (*Institutions*) $3.50 (*Individuals*)

Published for
the METAPHILOSOPHY FOUNDATION, INC. by
BASIL BLACKWELL
5 Alfred Street, Oxford, England

The Philosophical Forum

A Quarterly

VOLUME III WINTER 1971–1972 No. 2
Marjorie Grene *People and Other Animals*
Hugh Lehman *Approaches to the Explanation of Behavior*
Bernhard Rensch *Spinoza's Identity Theory and Modern Philosophy*
Janet L. Travis *Progressivism and the Human Supremacy Argument*
Hilde Hein *Comments on J. L. Travis, "Progressivism and the Human Supremacy Argument"*
Joseph P. DeMarco *The Concept of Race in the Social Thought of W. E. B. Du Bois*
Irving Thalberg *Justifications of Institutional Racism*
Joseph Agassi *The Twisting of the I. Q. Test*
Raymond Polin *Force and Its Political Uses in Hobbes*

VOLUME III SPRING, 1972 No. 3
SPECIAL ISSUE ON GYORGY LUKÁCS
The special issue will contain an article by Lukács previously unpublished in English, entitled *A Contribution to the Phenomenology of the Creative Mode.* Among the contributors are Andrew Arato, Andrew Feenberg, Ernst Fischer, Agnes Heller, György Márkus, Adam Schaff, and Willis H. Truitt.

VOLUME III SUMMER, 1972 No. 4
SPECIAL ISSUE ON ISLAMIC PHILOSOPHY

$10.00 a year for institutional subscribers; $8.00 a year for individual subscribers;
$5.00 a year for student subscribers, payable to
The Philosophical Forum
Department of Philosophy BOSTON UNIVERSITY
232 Bay State Road, Boston, Massachusetts 02215

THE REVIEW OF METAPHYSICS

A Philosophical Quarterly
Edited by Jude P. Dougherty

Volume XXV — Supplement in Honor of Paul Weiss on his 70th Birthday and the 25th Anniversary of the Founding of the *Review* — Spring 1972
Edited by Richard J. Bernstein

Andrew J. Reck, *Paul Weiss's Concept of Being*
Marc F. Griesbach, *The Epistemology of Paul Weiss*
Robert Neville, *Knowledge and Being: Comments on Griesbach and Reck*
Paul G. Kuntz, *The Achievement of Paul Weiss*
Kenneth L. Schmitz, *Art and Existence: Reflections on Paul Weiss's Modal Philosophy of Art*
Nathan Rotenstreich, *The Ontological and Epistemological Dimensions of History*
Charles Hartshorne, *Paul Weiss's* The God We Seek
Irwin C. Lieb, *Individuals and the Past*
Ellen S. Haring, *Actualities as Private and Public*
Paul Weiss, *A Response*
Bibliography of Paul Weiss's Publications

Individual Subscription $6.00 — Student Subscription $4.00
Institutional Subscription $10.00

The Catholic University of America, Washington, D.C. 20017
Twenty-Year Cumulative Index, 1947-67 is available for $3.00 a copy.

DISSERTATION ESSAY COMPETITION

The Publisher of *The Review of Metaphysics*, The Philosophy Education Society, Inc., announces the annual essay competition. The competition is open to participants who have been awarded the Ph.D degree in Philosophy in the United States or Canada during 1972. Essays must be a chapter from a dissertation, or a paper based directly upon the dissertation. Essays on any topic dealt with in the dissertation are acceptable. Essays submitted should not exceed twenty-five typewritten pages (double-spaced) and should be accompanied with a return postage envelope. The author of the prize-winning essay will receive $100 and it is expected that the essay will be published in *The Review of Metaphysics*. The best essay will be selected by the following committee: John J. Compton, Vanderbilt University: Donald Davidson, The Rockefeller University; Gerald J. Massey, University of Pittsburgh; and Jude P. Dougherty, the editor of *The Review of Metaphysics*, The Catholic University of America. Essays must bear a postmark no later than January 13, 1973 and should be sent to *The Review of Metaphysics*, The Catholic University of America, Washington, D. C. 20017. Essays should be marked as submissions for the competition. The winning essay will be announced by April 1, 1973. Additional inquiries concerning the competition may be directed to the above address.